

Searching for the Triple Higgs Boson Production and Beyond using Machine Learning at the ATLAS Experiment

Maggie Chen

Somerville College

University of Oxford



A thesis presented for the degree of

Doctor of Philosophy

2025

Dedicated to those I love and will fight to keep close.

Abstract

This thesis presents the first search for triple Higgs boson (HHH) production conducted at the Large Hadron Collider (LHC), using 126 fb^{-1} of proton-proton collision data at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ recorded by the ATLAS detector during Run-2 data-taking. The Standard Model (SM) HHH production is unique in its direct sensitivity to the trilinear and quartic Higgs self-coupling constants, therefore presenting an opportunity to detect deviations from their predicted values. These deviations are characterised by the Higgs self-coupling modifiers - κ_3 and κ_4 , and they have been simultaneously constrained for the first time at the LHC through the search for HHH production. This signature also provides an avenue to probe Beyond the Standard Model (BSM) theories in the extended Higgs sectors. Three models in particular are targeted, the "Two Real Singlet Model" (TRSM), "a simple model for dark matter and CP violation" (DM-CPV), and generic heavy scalars. The search is conducted in the final state of $b\bar{b}b\bar{b}b\bar{b}$, maximising statistics but at the cost of a large background arising from Quantum Chromodynamics (QCD) processes. The thesis focuses on the methodology developed to search for such a complex final state against this challenging background. The expected (observed) upper limit of the SM HHH signal strength is extracted to be 750 (760) times the predicted value at the 95% confidence level. Constraints on the Higgs self-coupling modifiers κ_3 and κ_4 are placed simultaneously by scanning over the profile likelihood ratio, assuming SM values for all other coupling constants. At the 95% confidence level, κ_3 is constrained between $-11 - 17$, and κ_4 between $-230 - 240$ for both the expected and observed limits, while assuming the SM value of the other modifier. For the targeted BSM models, the expected (observed) upper limits on the allowed cross-sections are placed between $46 - 350 \text{ fb}$ ($48 - 310 \text{ fb}$) for the TRSM and DM-CPV, and between $4.7 - 69 \text{ fb}$ ($5.7 - 38 \text{ fb}$) for generic heavy scalars with a narrow decay width, and between $5.2 - 53 \text{ fb}$ ($6.3 - 39 \text{ fb}$) for a wide decay width. This search, especially the simultaneous constraints on the coupling modifiers, sets the foundation for future HHH analyses. Beyond the main adopted method, several alternatives are explored to initiate discussions on new techniques for multi-Higgs analyses. A novel approach to search for new physics is presented, which utilises previously unexplored topologies within dataset-wide graph representations and Graph Neural Networks (GNNs). The method demonstrates the potential to surpass the performance of conventional machine learning methods in the case study of a BSM leptoquark search. The thesis also presents the feasibility of rejecting $b\bar{b}$ -jets arising from small-

angle $g \rightarrow b\bar{b}$ splitting, a common background in the ATLAS triggers, using machine learning. With the potential to reduced the readout rate, it could be highly relevant in future data-taking at high luminosity, and improve signal efficiencies for analyses targeting resolved b -quarks.

Acknowledgements

First and foremost, I would like to thank my supervisor Professor Todd Huffman, who has always been incredibly generous with his time and energy, and with whom I had some of the most enlightening discussions on physics. He has continuously supported, and at times constructively challenged, the many crazy ideas I proposed, which often led me to exploring exciting, unconventional avenues. I am also grateful for the guidance of Dr James Frost whose advice I value deeply, Dr Chris Pollard for patiently guiding me through the early stage of my PhD, and Dr Bill Balunas who led the $HHH \rightarrow 6b$ analysis in stride. My special gratitude is to Dr Holly Pacey, who has been the most amazing postdoc throughout the $HHH \rightarrow 6b$ analysis and the dataset-wide GNN project. Her insights on physics, machine learning, and career planning have been invaluable to me.

The highlight of my PhD had been my time at CERN. Geneva felt like home to me, and the people made it so. I was incredibly lucky to have Patrick as a flatmate, whom I see as family. My friends Sid and Anna, taught me to take all the simplicities and complications in life with grace; and Seb, has not only been an excellent collaborator at work but also a great partner in crime. My "Humboldt crew", Dan, TJ and Arely, shared my love for making mojitos and keeping them flowing through Dan's office window on the ground floor of Building 32. I was also honoured to serve as the social chair for the ATLAS flavour-tagging group, where I befriended many like-minded souls. Back in Oxford, Alessandro and Beth have been permanent sources of support in every aspect of my life. My adoption into Wolfson College by Beth and Will added an extra layer of joy and bartending skills to my repertoire. With Ynyr, I found my kindred spirit in adventure, who was always up for an impromptu trip to the Lamb and Flag to talk about nonsense or climbing up the Magdalen tower (from the outside) in formal dress. My new-found community, Gustavo and Sarah, made sure I was sane and well-caffeinated during thesis-writing.

Finally, I am grateful for my family, who supports me not through verbose gestures or daily care, but through their unwavering trust in my ability, strength and beliefs. Although continents apart, I always feel them with me, as they are a part of who I am.

Preface

It's all in the footnotes.

— Kevin Greif

During the last year of my PhD, I developed a casual interest in etymology, as a mode of procrastination that felt somewhat productive. The word thesis comes from the Greek word *thesis* (θέσις), meaning "a settling down", "a placing". It implies the purpose of a PhD is so that, by the end of it, one would be in a good position to settle on a conclusion or place a definitive statement on a subject. Indeed, in this thesis, I attempt to draw meaningful conclusions on some of the most interesting problems this mushy grey mass that sits atop my neck ever had the privilege to ponder. They include the feasibility of rejecting a background signature previously unaccounted for during data-taking in one of the largest particle physics experiments in the world, the ATLAS experiment (Chapter 3). Through a pioneering search for the triple Higgs boson production (Chapter 4), I contributed to the advancement on our knowledge about the nature of the electroweak symmetry breaking, the vacuum stability of the Universe, and the possibility of new physics beyond the highly successful Standard Model. Beyond these studies, there still remain many methods, theories, and conclusions that I cannot settle down on or place. Featured in this thesis are three alternative machine-learning-based methods to search for the triple Higgs boson production (Chapter 5), aiming to initiate constructive discussions on innovative methods for future iterations of the analysis. A novel approach for future searches of new physics at collider experiments is introduced, utilising previously unexplored features in dataset-wide graph representations (Chapter 6).

Indeed, the field of particle physics itself is no where near settling down either. At the Large Hadron Collider, the extensive physics programmes at its four major experiments are continuously looking for answers to some of the most fundamental question about the Universe. To this end, the experiments are expanding, with the planning of a Future Circular Collider coursing at full speed through small French village bureaucracies and the soft Molasse rocks beneath Lac Lemman in Geneva. There is much still to come.

Equally, on a personal level, I have never felt less settled as I near the end of my PhD journey. I am more curious than ever about a wider range of subjects, such as the implications of artificial intelligence on the epistemic process of scientific research, the potential applications of machine learning methods developed for particle physics in humanitarian efforts, the etymology of the word etymology¹... The journey, both for the field of particle physics and myself, will no doubt continue in the most exciting fashion.

Publications and contributions

The work presented in this thesis has been done by the author in collaboration with colleagues from institutes all over the world. The results and figures, unless stated otherwise, are made by the author or from the following publications which the author has contributed to:

Search for triple Higgs boson production in the $6b$ final state using pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector (ATLAS Collaboration, 2024 [1]).

The author was one of the main analysers contributing to the methodology development, trigger strategy optimisation and calibration, background estimation and statistical analysis. The work presented in Chapter 4 describes these efforts in detail.

Configuration, performance and commissioning of the ATLAS b -jet trigger for 2022 and 2023 LHC data-taking periods (ATLAS Collaboration, 2024 [2])

The author had been an active member of the ATLAS b -jet trigger group from 2022-2024, contributing to its algorithmic development, performance studies, calibration and operation. The development of the machine learning algorithm to reject bb -jets arising from small-angle $g \rightarrow b\bar{b}$ splitting presented in Chapter 3 is not included in this publication, but is archived in the trigger menu for potential deployment in the future, and published in [3]. The author contributed to the trigger rate monitoring and publication of the figures associated with the performance of the baseline b -tagging algorithm included in this publication.

¹The word comes from the Greek word "étymon" meaning "true sense", and the Greek suffix "-logia" meaning "the study of". Together, they mean "the study of the true meaning of words".

Table of Contents

1	Theoretical Background	11
1.1	The Standard Model of Particle Physics	11
1.2	Fundamental Forces	13
1.2.1	Quantum Electrodynamics (QED)	14
1.2.2	Quantum Chromodynamics (QCD)	15
1.2.3	Weak Interaction and Electroweak Unification	16
1.3	The Higgs Mechanism	18
1.4	The Standard Model Lagrangian	22
1.5	The Higgs Self-coupling	23
1.6	Single and Multi-Higgs Production	26
1.7	The Extended Higgs Sectors	29
2	Experiment	32
2.1	The Large Hadron Collider	32
2.2	The ATLAS Detector	34
2.2.1	The Inner Detector	35
2.2.2	The Electromagnetic Calorimeter (ECAL)	38
2.2.3	The Hadronic Calorimeter (HCAL)	38
2.2.4	The Muon Spectrometer	39
2.3	Energy, Luminosity and pile-up	40
2.4	The ATLAS Trigger System	41
2.4.1	Level-1 Trigger (L1)	42
2.4.2	High Level Trigger (HLT)	43
2.5	Physical Object Reconstruction	44
2.5.1	Tracks, Vertices and Calorimeter Topo-clusters	44
2.5.2	Jets	46
2.5.3	Muons	52

3	Rejecting $g \rightarrow b\bar{b}$ Splitting in the ATLAS b-jet Trigger	53
3.1	The ATLAS b -jet Trigger	53
3.2	b -tagging Algorithms	54
3.3	$g \rightarrow b\bar{b}$ Rejection Algorithm	55
3.3.1	Training Setup	58
3.3.2	Input Variables	59
3.3.3	Performance	62
3.4	Working Point Optimisation	62
3.4.1	Working Point Scan	64
3.4.2	2-dimensional Discriminant Visualisation	65
3.5	Implication in the b -jet Trigger	66
3.6	Conclusions and Outlooks	69
4	The Search for Triple Higgs Boson Production in the 6-b-quark Final State	70
4.1	Introduction	70
4.1.1	Physics Interpretations	72
4.1.2	Analysis Strategy Overview	74
4.2	Data and Monte Carlo Samples	75
4.3	Event Reconstruction and Selection	77
4.3.1	Trigger	77
4.3.2	Object Definitions	77
4.3.3	Event Selection and Categorisation	79
4.3.4	Jet Pairing	80
4.4	Signal and Background Discrimination	81
4.4.1	Training Setup	81
4.4.2	Input Feature Selection	85
4.4.3	Performance	93
4.5	Background Estimation	93
4.5.1	Estimation Method	96
4.5.2	Validation	100

4.6	Systematic Uncertainties	100
4.6.1	Detector Uncertainties	102
4.6.2	Theory Uncertainties	103
4.6.3	Background Estimation Uncertainties	104
4.7	Jet-level Trigger Calibration	107
4.7.1	Simulated Samples and Event Selection	107
4.7.2	Scale Factor Derivation Procedure	108
4.7.3	Calibration Results	110
4.7.4	Systematic Uncertainties	113
4.7.5	b -jet level Trigger Uncertainty	113
4.8	Statistical Analysis	117
4.8.1	The Likelihood Function and p -value	118
4.8.2	The Test Statistics	119
4.8.3	The CLs Method	120
4.8.4	Fit Setup	121
4.9	Results	123
4.9.1	HHH Production Cross-section Upper Limits	123
4.9.2	Upper Limits on κ_3 and κ_4	126
4.10	Conclusions	131
4.11	Outlooks	133

5 Alternative Strategies for $HHH \rightarrow 6b$ 135

5.1	Event Categorisation in the m_H Volume	135
5.1.1	Analysis Regions	136
5.1.2	Alternative Jet Pairing Methods	139
5.1.3	Background Reduction Inside the Signal Region	140
5.1.4	Conclusions	143
5.2	Parametrised Neural Network	145
5.2.1	Training	146
5.2.2	Performance Compared to the Baseline Model	147

5.2.3	Parametrisation Schemes	147
5.2.4	Conclusions	149
5.3	Kinematic Regression	150
5.3.1	Deep Sets Architecture	151
5.3.2	Training	152
5.3.3	Performance Comparison with Baseline	152
5.3.4	Conclusion	154
6	Dataset-wide Graph Neural Networks for Searches of New Physics	156
6.1	Introduction	156
6.2	Theoretical Background	159
6.2.1	Graph Construction	159
6.2.2	Distance Metrics	161
6.2.3	Graph Neural Network	162
6.2.4	Graph Convolution Functions	165
6.3	Case Study: Searching for Leptoquarks	168
6.3.1	Simulated Events and Event Selection	168
6.3.2	Event Similarity in Kinematic Space	170
6.3.3	Event Similarity in Latent Space	175
6.3.4	Kinematic Convolution	179
6.4	Model Scalability	180
6.5	Model Implementation and Training	182
6.5.1	Transductive Learning	184
6.6	Model Performance	186
6.6.1	Area Under the ROC Curve (AUC)	186
6.6.2	Z Significance	187
6.7	Conclusions	189
6.8	Future Studies	194
7	Conclusions	197

References	201
Appendix A DL1dbb Input Variables	214
Appendix B Event Display of a $HHH \rightarrow 6b$ Candidate Event	217
Appendix C Additional Information on ResDNN, NonresDNN and HeavyresDNN	218
Appendix D HHH Signal Contamination in Various b-tag Regions	226
Appendix E Shape Variations of the HHH Background Estimate	228
Appendix F EmbeddingNet Algorithm	230
F.1 Parameter Scan	230
F.2 Training	230
F.3 Embedding Features in Latent Space	231
Appendix G Additional Information on the Dataset-wide GNNs	233
G.1 Kinematic Convolutions	233
G.2 GNN Training	237
Appendix H A Quick Study on My Thesis Writing Habits	240

1 | Theoretical Background

The quality of a cappuccino at CERN varies as a function of time and space.

— Ynyr Harris

One of mankind's greatest endeavours is to describe and understand observed phenomena in the Universe. The Standard Model (SM) of particle physics is the most accurate theoretical framework to date in describing the observed matter as elementary particles and the interactions between them. Its success is revealed through the agreement between theoretical predictions and experimental data gathered at various particle physics experiments around the world. However, there remain many open questions, leading to a plethora of extensions to this theory. Many of them are associated with the Higgs boson, an important particle in the SM. These in particular give rise to the extended Higgs sectors, acting as a playground for theorists and experimentalists to search for plausible Beyond the Standard Model (BSM) physics. This chapter first gives an overview of the SM as a gauge theory, covering the particles and three out of the four fundamental interactions it describes. Focusing on the Higgs boson, Section 1.3 outlines the Higgs mechanism, a mechanism of spontaneously breaking the electroweak symmetry. Section 1.3 and Section 1.5 focus on the Higgs mechanism, the Higgs energy potential and its relation to the Higgs self-coupling, both of which are of particular relevance to the work presented in this thesis. Finally, an overview of the theories in the extended Higgs sectors that seek to solve some of the shortcomings of the SM is given in Section 1.7.

1.1 The Standard Model of Particle Physics

The Standard Model is a quantum field theory, which describes fundamental particles as excitations of the quantum fields that permeate spacetime. The dynamics of these quantum fields are described by the Lagrangian of the system, and are governed by the principle of least action. The Lagrangian is invariant under certain gauge transformations such that the physical system remains independent of the chosen gauge. Its invariance under symmetry transformations at each point in spacetime is referred to as local gauge invariance. The symmetry groups which arise from the transformations

between allowed gauges in the SM are

$$SU(3)_c \times SU(2)_L \times U(1)_Y \quad (1.1)$$

These groups introduce the gauge fields to ensure the local gauge invariance of the Lagrangian, which correspond to the fundamental forces. The strong interaction (QCD) is described by $SU(3)_c$ symmetry, and the electroweak interaction is described by $SU(2)_L \times U(1)_Y$ symmetries. The excitations of these gauge fields give rise to the gauge bosons that mediate the forces. The subscripts here relate to the charges or the properties unique to each force. The subscript c represent the colour charge in the strong interaction, L represents left-handedness of the particles that the weak force acts upon, and Y corresponds to the hypercharge related to the electromagnetic interaction after spontaneous symmetry breaking.

Below the vacuum expectation energy of 246 GeV, the electroweak force breaks down into the weak interaction and electromagnetism (QED). This breaking of symmetry is due to the Higgs mechanism which gives rise to massive gauge bosons. The particles within the theory are categorised by their spin values - particles with half-integer spins are called *fermions*, and particles with integer spins are called *bosons*. Fermions are particles that obey the Fermi-Dirac statistics and constitute the known matter in the Universe, and bosons are the force-carrying particles that mediate the interactions between them. Fermions are further divided into quarks and leptons, and they are arranged in three generations of increasing masses, summarised in Table 1.1. All fermions interact via the weak force, and the ones carrying electric charges also interact via the electromagnetic force. Ones that do not carry electromagnetic forces are called neutrinos. Quarks differ from leptons as they also carry colour charges, and therefore interact via the strong force. All fermions have anti-matter counterparts, which have the same masses but the opposite electric charge and other quantum numbers.

Fermions interact through the exchange of gauge bosons, represented as vector fields with integer spin values. They are excitations in gauge vector fields that are invariant under gauge symmetries. The interactions each of them mediates are summarised in Table 1.2. Note that gravity is not described by the SM, and the force-carrying particle - the "graviton" - is only a hypothetical particle

Type	Generation			Spin	Electric Charge
	1st	2nd	3rd		
Quarks	u	c	t	1/2	+2/3
	d	s	b	1/2	-1/3
Leptons	ν_e	ν_μ	ν_τ	1/2	0
	e	μ	τ	1/2	-1

Table 1.1: Table summarising the fermions within the Standard Model and their corresponding spins and electric charges. They are divided into quarks and leptons and arranged into three generations of increasing mass from left to right. The top row of the quarks are referred to as up-type quarks, and the bottom row as the down-type quarks. Similarly

that does not exist in the SM or predicted by a complete quantum field theory.

Fundamental force	Gauge boson(s)
Electromagnetism	Photon (γ)
Weak force	W boson ($W^\pm$), Z boson (Z)
Strong force	Gluon (g)

Table 1.2: Table summarising the gauge bosons and the fundamental force they mediate.

The last particle described by the SM is the Higgs boson, which is a neutral scalar boson of spin-0. It is the excitation of the Higgs field, a field that enables the electroweak symmetry breaking, discussed in more detail in Section 1.3. This mechanism gives rise to a non-zero vacuum expectation value (vev), the masses of the W and Z bosons, and through Yukawa couplings the masses of the fermions. The Lagrangian of the Higgs field also predicts the interactions between multiple Higgs bosons, referred to as the Higgs self-coupling, discussed with more detail in Section 1.5.

1.2 Fundamental Forces

The behaviours of fermions are described by the Dirac Equation:

$$(i\not{\partial} - m)\Psi = 0 \quad (1.2)$$

where $\not{\partial} = \gamma_\mu \partial^\mu$, ∂^μ is the contravariant derivative operator with respect to spacetime coordinates and γ_μ are the Dirac matrices. m is the mass of the fermion, and Ψ is the 4-component Dirac particle spinor. The negative energy solutions to the Dirac equation also directly predict the existence of anti-particles. Following it, the Lagrangian of a freely propagating fermionic field Ψ that is Lorentz invariant can be expressed as

$$\mathcal{L} = \bar{\Psi}(i\not{\partial} - m)\Psi \quad (1.3)$$

where $\bar{\Psi} = \Psi^\dagger \gamma^0$. The behaviours of the bosons are described by vector fields with integer spins.

1.2.1 Quantum Electrodynamics (QED)

QED describes the interaction between particles that carry electromagnetic charges mediated by photons. The Dirac spinor fields under the $U(1)$ symmetry transforms as

$$\Psi(x) \rightarrow \Psi' = e^{i\theta}\Psi \quad (1.4)$$

where θ is a global phase. If the phase invariance is local, that is if θ depends on space-time points x , $\theta \rightarrow \theta(x)$ and additional terms in the gauge vector field A_μ are needed for the Lagrangian to remain invariant under this transformation

$$A_\mu \rightarrow A'_\mu = A_\mu + \frac{1}{g}\partial_\mu\theta(x) \quad (1.5)$$

Requiring it to be invariant under Lorentz transformation, the Lagrangian of the gauge vector field can be expressed as

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad (1.6)$$

where $F_{\mu\nu}$ is the strength of the gauge vector field $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$. With the introduction of this vector field A_μ and solving the Dirac equation, the Lagrangian for QED can be written as

$$\mathcal{L}_{\text{QED}} = -\bar{\Psi}(\gamma^\mu \partial_\mu + m)\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + ieA_\mu \bar{\Psi}\gamma^\mu\Psi \quad (1.7)$$

Breaking this down, the first term describes propagation of the fermionic field Ψ , which is the charged particle. The second term describes the propagation of the gauge vector field A_μ , that is the photon. The third term describes the interaction between these two fields. Overall, it describes an interaction between Dirac fermions, such as the annihilation of an electron and positron via a QED interaction vertex, shown in Figure 1.1.

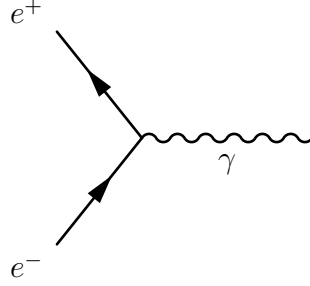


Figure 1.1: Feynman diagram for the annihilation of an electron and a positron via an interaction vertex for the QED Lagrangian.

1.2.2 Quantum Chromodynamics (QCD)

QCD is a quantum field theory of the strong interaction between quarks, which carry colour charges. It is mediated by the gluon. The dynamics of the strong interaction are described by the terms in the Lagrangian

$$\begin{aligned}\mathcal{L}_{QCD} &= \sum_q i g_s \bar{q}_i \gamma^\mu (G_\mu^a T^a)_{ij} q_j \\ &= \sum_q i \frac{g_s}{2} \bar{q}_i \gamma^\mu (G_\mu^a \gamma_a)_{ij} q_j\end{aligned}\tag{1.8}$$

where g_s is the strong coupling constant. The sum is over all quark flavours q , and i runs over the colour state. T^a are the 8 generators of the $SU(3)$ symmetry group which can be expressed using Gell-Mann matrices λ^a . The Gell-Mann matrices are of dimension (3×3) , leading to the three possible colour states of the quarks, red (r), green (g) and blue (b). The 8 generators of the $SU(3)$ symmetry group corresponds to the 8 colour states of the gluon: $r\bar{g}$, $r\bar{b}$, $g\bar{r}$, $g\bar{b}$, $b\bar{r}$, $b\bar{g}$, $\frac{1}{\sqrt{2}}(r\bar{r} - b\bar{b})$ and $\frac{1}{\sqrt{6}}(r\bar{r} + b\bar{b} - g\bar{g})$. The gluon fields $G_{\mu\nu}^a$ transform as

$$G_{\mu\nu}^a \rightarrow G_{\mu\nu}'^a = \partial_\mu G_\nu^a - \partial_\nu G_\mu^a + g_s f^{abc} G_\mu^b G_\nu^c\tag{1.9}$$

where f^{abc} are the structure constants of the $SU(3)$ group, a result of the non-commuting generators. This additional term also gives rise to the self-interaction between the gluon fields in the last term, shown in Figure 1.2, reflecting the non-abelian nature of $SU(3)$. This leads to asymptotic freedom, where the strong coupling g_s is weaker at higher energies (shorter distances) and stronger at low energies (longer distances), resulting in the confinement of quarks inside hadrons. Figure 1.2 shows examples of QCD interactions between quarks and gluons, and of gluon self-interactions that include the interaction vertex for the QCD Lagrangian.

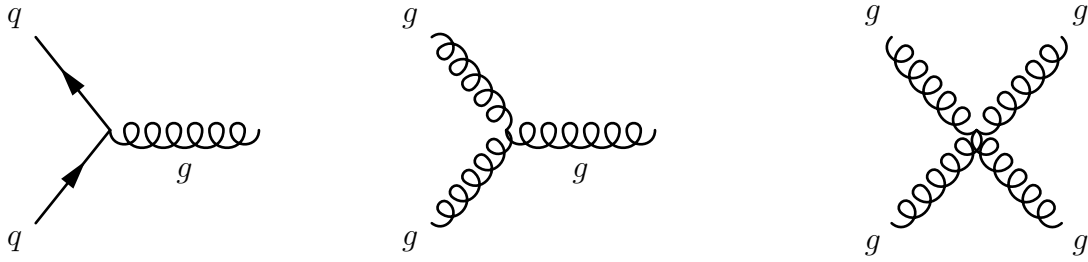


Figure 1.2: Feynman diagrams of QCD interactions between quarks and gluons and gluon self-interactions.

1.2.3 Weak Interaction and Electroweak Unification

The weak interaction can change the flavours of the fermions. It is mediated by massive gauge bosons, the charged $W^\pm$ bosons and the neutral Z boson. The charged-current weak interaction mediated by the $W^\pm$ bosons therefore couple fermions that differ by one unit of electric charge, such as the interaction in Figure 1.3. It is the only interaction that does not conserve parity in the SM. The parity violation is observed experimentally in the β -decay of the polarised Cobalt-60, where there is a preferred direction of the emitted electrons in a magnetic field. This confirms that the weak interaction does not have a 4-vector current of the form $j^\mu = \bar{u}(p')\gamma^\mu u(p)$. Instead, it takes the form of a linear combination of vector and axial vector currents ($V - A$). For example for the exchange of a W boson, it is given by

$$j^\mu = \frac{g_W}{\sqrt{2}} \bar{u}(p') \frac{1}{2} \gamma^\mu (1 - \gamma^5) u(p) \quad (1.10)$$

The charged-current weak interaction only couples to particles with left-handed (LH) chirality and anti-particles with right-handed (RH) chirality. To demonstrate this, one can decompose a particle

spinor into the LH and RH chiral components:

$$\begin{aligned}
 u &= \frac{1}{2}(1 + \gamma^5)u + \frac{1}{2}(1 - \gamma^5)u \\
 &= P_R u + P_L u
 \end{aligned} \tag{1.11}$$

The $V - A$ vertex factor of the weak interaction already includes the LH chiral component. The contribution to the 4-vector current from both RH chiral states of the particle spinor is zero, and for the anti-particle spinor the LH chiral states' contribution is zero. Therefore only the LH chiral particle states and the RH anti-particle states participate in the charged-current weak interaction. Experimentally, this is observed in the charged pion decay. Due to the small mass of the neutrinos produced, its helicity¹ and chiral states can be treated as equivalent. Therefore the antineutrinos from a π^- decay are always RH. To conserve angular momentum, the charged leptons are also produced in the RH helicity state. However, the helicity and chiral states of the leptons are not equivalent due to their masses, therefore they can have a LH chiral states whilst having a RH helicity state, still allowing the weak interaction to occur. However, the weak decay to a RH helicity particle state is highly suppressed, by a factor that is inversely proportional to the mass of the lepton. Experimentally, this is observed as the highly suppressed pion decays to electrons compared to muons.

The strength of the weak interaction can be determined by measuring the lifetimes of weakly decaying particles, such as muons, where the lifetime is inversely proportional to the coupling strength. It was found to be much weaker than the strong and electromagnetic interactions. The mass eigenstates and the weak eigenstates of the quarks are not equivalent, and the mixing of these two eigenstates is described by the Cabibbo-Kobayashi-Maskawa (CKM) matrix:

$$\begin{bmatrix} d' \\ s' \\ b' \end{bmatrix} = \begin{bmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{bmatrix} \begin{bmatrix} d \\ s \\ b \end{bmatrix} \tag{1.12}$$

¹Helicity is the projection of the spin of the particle onto its momentum. It is different from chirality in the sense that it is dependent on the frame of the particle's motion.

where the weak eigenstates (d', s', b') are related to the mass eigenstates (d, s, b) by the CKM matrix V_{CKM} . The elements of this matrix indicate the probabilities of the change in quark flavour across the three quark generations.

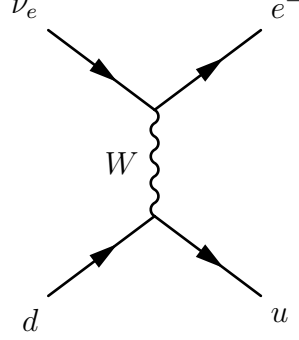


Figure 1.3: Feynman diagram of a charged-current weak interaction between leptons and quarks via the exchange of a W boson.

In addition to the weak charged currents, the $SU(2)_L$ gauge symmetry suggests the existence of a weak neutral current. Experimentally, this was discovered by neutrino scattering experiments, for example via an interaction shown in Figure 1.4. It was shown that this neutral current coupled to both LH and RH chiral states, albeit unequally. The need for a gauge theory that describes the charge-current weak interaction and its similarity with electromagnetism led to the electroweak unification. The electroweak model replaces the $U(1)$ gauge symmetry of electromagnetism with a new local gauge symmetry $U(1)_Y$. This results in a gauge field B_μ that couples to a new type of charge, the weak hypercharge $Y = 2(Q - I_W^{(3)})$, where Q is the electric charge and $I_W^{(3)}$ is the third component of the weak isospin, the quantum number associated with the weak interaction. The neutral fields of $U(1)_Y$ are mixed with the $SU(2)_L$ gauge symmetry, described by the mixing angle θ_W . The resulting gauge fields of the $SU(2)_L \times U(1)_Y$ gauge symmetry group correspond to four massless bosons $W^\pm, Z^0, \gamma$. However, experimental evidence shows that the $W^\pm$ and Z^0 bosons are massive, which points to an additional mechanism that gives rise to their masses.

1.3 The Higgs Mechanism

The spontaneous breaking of the electroweak symmetry is the mechanism that gives masses to the gauge bosons without breaking the actual gauge symmetry. This is known as the Higgs mechanism,

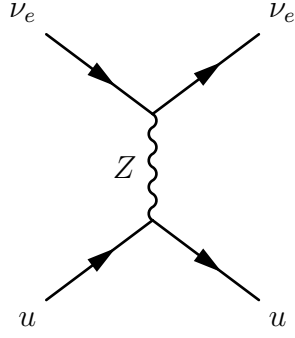


Figure 1.4: Feynman diagram of a neutral-current weak interaction between leptons and quarks via the exchange of a Z boson, such as one observed in a neutrino scattering experiment.

a crucial part of the SM. It introduces a spin-0 scalar boson, the *Higgs boson*, and the Higgs field with a non-zero vacuum expectation value (vev). The simplest Higgs model describes a weak isospin doublet ϕ (the Higgs doublet) consisting of two complex scalar fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \quad (1.13)$$

where one of the scalar fields ϕ^0 is neutral in order to generate the masses of the neutral electroweak gauge bosons, and ϕ^+ is charged to give the longitudinal degrees of freedom to generate the masses of the charged W bosons. The behaviours of these scalar fields are governed by the Lagrangian

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - V(\phi) \quad (1.14)$$

where $V(\phi)$ is the Higgs potential, expressed as

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (1.15)$$

The vacuum state is defined as the minimum of the potential $V(\phi)$, which corresponds to the lowest energy state of the field ϕ . For $\mu^2 < 0$, there exists an infinite number of non-zero minima which satisfies $\langle \phi \rangle = \frac{\nu^2}{2} = -\frac{\mu^2}{2\lambda}$, where ν is the vev of the Higgs field. Below a certain energy, the minimum of the potential must yield a non-zero vev only for the neutral scalar field ϕ^0 in order for the neutral

photon to remain massless. Expanding around the vev, the fields can be expressed as

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1(x) + i\phi_2(x) \\ \nu + \eta(x) + i\zeta(x) \end{pmatrix} \quad (1.16)$$

A massive scalar particle and three massive Goldstone bosons result from the spontaneous symmetry breaking. In order to eliminate the unphysical coupling between the massive scalar and the Goldstone bosons, as a spin-1 gauge field cannot transform into a spin-0 scalar field, a unitary gauge is applied. As a result, the three Goldstone bosons are effectively "gauged-away" and the Higgs doublet field is simplified to

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ \nu + h(x) \end{pmatrix} \quad (1.17)$$

Returning to the expression of the Lagrangian in Equation 1.14, the derivatives need to be replaced with the appropriate covariant derivatives such that the $SU(2)_L \times U(1)_Y$ local gauge symmetry is respected:

$$\partial_\mu \rightarrow D_\mu = \partial_\mu + ig_W \mathbf{T} \cdot \mathbf{W}_\mu + ig' \frac{Y}{2} B_\mu \quad (1.18)$$

where $\mathbf{T}$ are the generators of the $SU(2)$ symmetry, $\mathbf{W}_\mu$ correspond to the three weak isospin fields, B_μ is the weak hypercharge field and Y the weak hypercharge as introduced before. The Lagrangian can thus be written as

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (1.19)$$

where the term $(D_\mu \phi)^\dagger (D^\mu \phi)$ generates the masses of the gauge bosons. Expanding it using the expression for the covariant derivative in Equation 1.18, the terms in the Lagrangian quadratic in the gauge boson fields are

$$\underbrace{\frac{1}{8} \nu^2 g_W^2 (W_\mu^{(1)} W^{(1)\mu} + W_\mu^{(2)} W^{(2)\mu})}_{\text{Masses of the charged gauge bosons}} + \underbrace{\frac{1}{8} (g_W W_\mu^{(3)} - g' B_\mu) (g_W W^{(3)\mu} - g' B^\mu)}_{\text{Masses of the neutral gauge bosons}} \quad (1.20)$$

For the charged W^+ and W^- bosons, their mass terms appear as $\frac{1}{2}m_W^2 W_\mu^{(1)} W^{(1)\mu}$ and $\frac{1}{2}m_W^2 W_\mu^{(2)} W^{(2)\mu}$. The mass of the W boson can therefore be identified as $m_W = \frac{1}{2}g_W\nu$. The terms quadratic in the neutral $W^{(3)}$ field can be written as

$$\frac{\nu^2}{8} \begin{pmatrix} W_\mu^{(3)} B_\mu \end{pmatrix} \underbrace{\begin{pmatrix} g_W^2 & -g_W g' \\ -g_W g' & g'^2 \end{pmatrix}}_{\text{Mass matrix}} \begin{pmatrix} W^{(3)\mu} \\ B^\mu \end{pmatrix} \quad (1.21)$$

Diagonalising the mass matrix yields two eigenvalues, one being $m_A = 0$ which corresponds to the massless photon, and the other is $m_Z = \frac{1}{2}\nu\sqrt{g_W^2 + g'^2}$ corresponding to the mass of the Z boson. Expressing the ratio of the couplings of the $U(1)_Y$ and $SU(2)_L$ symmetries in the form of $\frac{g'}{g_W} = \tan \theta_W$, the photon and the Z fields can be written as the linear combinations of the two fundamental fields

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B^\mu \\ W^{(3)\mu} \end{pmatrix} \quad (1.22)$$

Using the same ratio, the mass of the Z boson can be expressed as $m_Z = \frac{1}{2} \frac{g_W}{\cos \theta_W} \nu$, which leads to an interesting relation to the mass of the W boson: $\frac{m_W}{m_Z} = \cos \theta_W$. This relation can be experimentally verified, providing evidence for the Higgs mechanism.

For fermions masses, the mass term in the Dirac Lagrangian does not respect the $SU(2)_L \times U(1)_Y$ gauge symmetry, due to the fact that the LH and RH chiral states transform differently. Therefore they cannot be directly added to the Lagrangian. However, the combinations of $\bar{L}\phi R$ and $\bar{R}\phi L$ are invariant under both $SU(2)_L$ and $U(1)_Y$ gauge transformations, and so are their Hermitian conjugates. Therefore, a term in the form of $-g_f(\bar{L}\phi R + \bar{R}\phi^\dagger L)$ can be added to the Lagrangian as it satisfies the necessary gauge symmetry of the SM. For example, after spontaneous symmetry breaking and in the unitary gauge, the mass term corresponding to the $SU(2)_L$ doublet containing

the electron can be expressed as

$$\begin{aligned}\mathcal{L}_e &= -\frac{g_e}{\sqrt{2}}\nu(\bar{e}_L e_R + \bar{e}_R e_L) - \frac{g_e}{\sqrt{2}}h(\bar{e}_L e_R + \bar{e}_R e_L) \\ &= -m_e \bar{e}e - \frac{m_e}{\nu}\bar{e}e h\end{aligned}\tag{1.23}$$

where g_e is the Yukawa coupling of the electron to the Higgs field. Although not predicted by the Higgs mechanism, its value is chosen as $g_e = \sqrt{2}\frac{m_e}{\nu}$ such that the mass of the electron matches the observed value. The first term in Equation 1.23 corresponds to the coupling of the electron to the Higgs field through its non-zero vev, which gives rise to the electron mass. The second term describes the coupling between the electron and the Higgs boson. However, only the masses of the fermions in the lower component of the $SU(2)_L$ doublets, namely the leptons and down-type quarks, can be generated this way. This is due to the non-zero vev in the lower component of the Higgs doublet ϕ^0 . Neglecting neutrinos due to their small masses, the mass terms for the up-type quarks can be generated using the combinations $\bar{L}\phi_c R + \bar{R}\phi_c^\dagger L$, where ϕ_c is the conjugate doublet of the fields in Equation 1.17. The formalism of their masses then follows exactly as before in the example for the electron.

An interesting consequence of this formalism is that the Yukawa coupling of the top quark with a mass of $m_t \sim 173.5 \pm 1.0$ GeV is almost exactly 1. Another consequence is that the small values of the neutrino masses result in their Yukawa couplings to be on the orders of 10^{-12} , which differs from other fermions. This suggests that the neutrino masses may be generated via other mechanisms.

1.4 The Standard Model Lagrangian

With all the components of the SM at hand, the most general SM Lagrangian that is renormalisable, obeying the $SU(3) \times SU(2) \times U(1)$ symmetry group, can be thus formulated:

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_{fermion} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}\tag{1.24}$$

where components individually are

$$\mathcal{L}_{gauge} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}W_{\mu\nu}^i W^{\mu\nu i} - \frac{1}{4}G_{\mu\nu}^a G^{\mu\nu a} \quad (1.25)$$

$$\mathcal{L}_{fermion} = \sum_f \bar{L}_f i\gamma^\mu D_\mu L_f + \sum_f \bar{R}_f i\gamma^\mu D_\mu R_f + h.c. \quad (1.26)$$

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (1.27)$$

$$\mathcal{L}_{Yukawa} = \sum_{f \in \{e, \mu, \tau, d, s, t\}} g_f \bar{L}_f \phi R_f - \sum_{f \in \{u, c, t\}} g_f \bar{L}_f \phi_c R_f + h.c. \quad (1.28)$$

$\mathcal{L}_{gauge}$ describes the dynamics of the gauge boson fields, and $\mathcal{L}_{fermion}$ describes the dynamics of the fermions. $\mathcal{L}_{Higgs}$ encapsulates the kinematic and mass terms of the Higgs boson, its interaction with other gauge fields and with itself. Finally, $\mathcal{L}_{Yukawa}$ captures the mass terms of the fermions except for the neutrinos and their interactions with the Higgs field and the Higgs boson.

Within the SM, there exists 19 free parameters: 3 Yukawa couplings in the lepton sector and 6 in the quark sector, 3 CKM matrix mixing angles and 1 CP violating phase, 3 gauge couplings of the vector fields, the Higgs mass parameter μ^2 and the Higgs self-coupling λ , and the strong CP phase θ_{QCD} . They are to be determined experimentally for the theory to be fully determined.

The success of the SM theory is demonstrated repeatedly via various measurements of its processes made by numerous experiments. The measurements of the cross-sections of various SM processes at different centre-of-mass energies are summarised in Figure 1.5.

1.5 The Higgs Self-coupling

The Higgs self-coupling is an interaction predicted by the Higgs component of the SM Lagrangian. Expanding the Higgs potential around the vev, the Lagrangian component $\mathcal{L}_{Higgs}$ includes the terms

$$-V_0 - \lambda \nu^2 h^2 - \lambda \nu h^3 - \frac{1}{4} \lambda h^4 \quad (1.29)$$

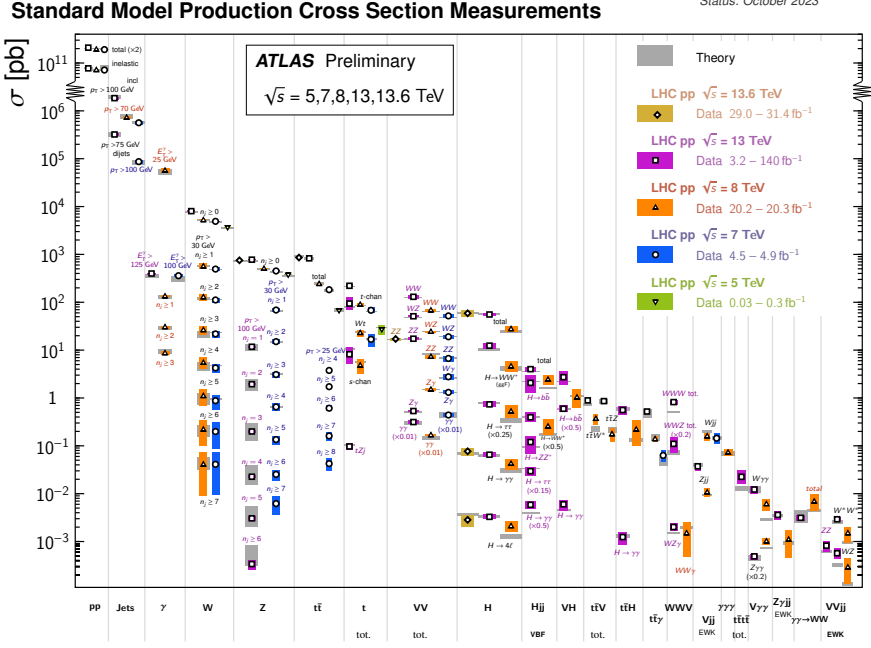


Figure 1.5: The total and fiducial cross-sections of numerous SM processes and the comparison to the corresponding theoretical predictions. The figure is taken from [4]

The term $\lambda_\nu h^3$ describes a three-point interaction vertex, with a coupling strength of $\lambda_3 = \frac{3m_H^2}{\nu}$ that corresponds to the *trilinear Higgs self-coupling*. The term $\frac{1}{4}\lambda h^4$ describes a four-point interaction vertex, with a coupling strength of $\lambda_4 = \frac{3m_H^2}{\nu^2}$ that corresponds to the *quartic Higgs self-coupling*. The Feynman diagrams for these Higgs self-interactions are shown in Figure 1.6. The production of

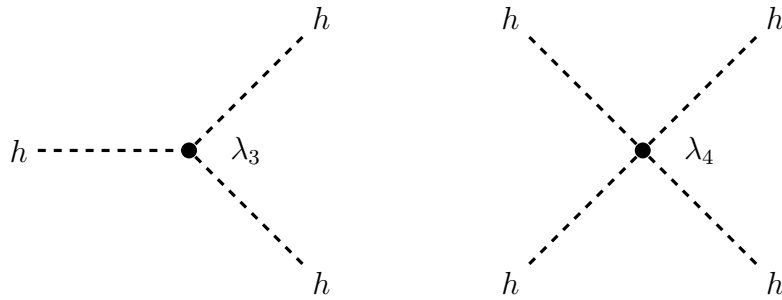


Figure 1.6: Feynman diagrams for the three-point and four-point self-interactions of the Higgs field, where λ_3 and λ_4 are the strengths of the respective interactions.

multiple Higgs bosons (multi-Higgs production) is sensitive to the strength of these self-couplings, therefore providing an experimental avenue to determine their values. Any deviation from the SM prediction can be characterised by the coupling modifiers κ_3 and κ_4 , defined as the ratio between the measured values and their SM predictions. They determine the shape of the Higgs potential, and have important implications on the phase transition of the early Universe.

About 10^{-10} seconds after the Big Bang, a phase transition of the Universe occurs which separates the fundamental forces and breaks $SU(2)_L \times U(1)_Y$ symmetry. Before this transition, the temperature is above a critical point T_C , where the electroweak force is unified and the vacuum state occurs at $h = 0$. As the Universe cools and temperature drops below T_C , the electroweak symmetry becomes broken and the vev non-zero. This change of the physical state of the Universe as a system is referred to as the electroweak phase transition (EWPT), which could occur via a violent first-order phase transition, or a gentle second-order phase transition. A second-order transition implies a smooth breaking of electroweak symmetry where the vev of the Universe sits in a global minimum. A first-order transition however would result in a local minimum of the effective Higgs potential at $\text{vev} = 0$, and global minima at a non-zero value. This creates a barrier between the two minima such as in Figure 1.7, and causes parts of the Universe to tunnel to the global minima, which manifests as bubbles containing the broken phase expanding into the cosmological plasma with symmetric phase [5]. These bubbles rapidly expand, collide and coalesce until the broken phase fills the entire plasma. During this process, particles in the cosmological plasma scatter off the bubble wall and generate asymmetries in baryon densities in front of it. They in turn diffuse into the plasma, generating the matter anti-matter asymmetry observed today. This process is referred to as *Electroweak Baryogenesis*, which requires a strong first-order EWPT, and is not possible with a second-order EWPT as it does not generate the non-equilibrium phase required. In the perturbative analysis of the EWPT, a first-order transition can be allowed by the effective Higgs potential depending on both the trilinear and quartic Higgs self-couplings [6]. Determining these couplings experimentally is therefore crucial in understanding the nature of the EWPT and the matter-antimatter asymmetry in the Universe. An array of BSM models in the extended Higgs sectors detailed in Section 1.7 also predicts a strong first-order EWPT, and are correlated with a large deviation in the Higgs self-coupling constants from their SM values. Furthermore, the shape of the Higgs potential also depends on these coupling constants. Any additional minima in the shape point to the metastability of the Universe which could result in the tunnelling to a new minimum in the potential. A strong first-order EWPT could be probed through astrophysical observations or particle physics experiments. The former attempts to measure gravitational waves created from the

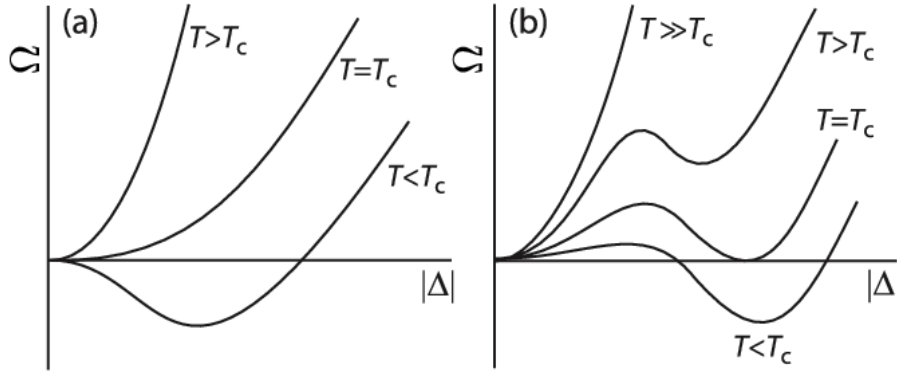


Figure 1.7: The thermodynamic potential $\Omega(|\Delta|)$ as a function of the order parameter Δ . In the case of EWPT, the thermodynamic potential here is analogous to the effective Higgs potential, and the order parameter is the vev. (a) shows the behaviour of the potential undergoing a second-order phase transition at different temperatures, and (b) shows the corresponding for a first-order phase transition. The figure is taken from [7].

phase transition at LISA [8]. The latter probes through searching for multi-Higgs productions at collider experiments. The discovery of the Higgs boson in 2012 by the ATLAS [9] and CMS [10] experiments seeded the extensive physics programme in measuring its properties, including its self-interaction. At the current centre-of-mass energy operated at the Large Hadron Collider (LHC), the production of single and double Higgs bosons are within experimental reach. The searches for the pair production of the Higgs boson conducted at the LHC place increasingly stringent constraints on λ_3 .

1.6 Single and Multi-Higgs Production

The Higgs self-coupling can be probed through single and multi-Higgs productions. An extensive search programme for these processes is well-established at the ATLAS experiment.

The trilinear Higgs self-coupling λ_3 can be probed through the next-to-leading order correction in single Higgs production as well as di-Higgs production. At $\sqrt{s} = 13$ TeV, the leading di-Higgs production modes are gluon-gluon fusion (ggF) with a cross-section of 31.05 fb at next-to-next-to leading order (NNLO) [11], and vector boson fusion (VBF) with a cross-section of 1.73 fb at the next-to-next-to-next-to leading order (N³LO) [12]. Figure 1.8 shows the Feynman diagrams for

these production modes that rely on the trilinear Higgs self-coupling constant λ_3 . The ATLAS experiment has conducted searches for these productions spanning a wide variety of final states, many of which exploit the dominant decay mode of $H \rightarrow b\bar{b}$, while others take advantage of the reduced background associated with decays to leptons or photons but trading off a reduced branching ratio. A search in the $b\bar{b}b\bar{b}$ final state using the full Run-2 dataset [13] constrained the value of κ_3 to be between -3.5 and 11.3 at 95% confidence level. Combining results from the full Run-2 dataset of the three most sensitive final states, namely $b\bar{b}b\bar{b}$, $b\bar{b}\tau^+\tau^-$ and $b\bar{b}\gamma\gamma$, more stringent constraints on κ_3 are placed to be between -1.2 and 7.2 at 95% confidence level [14].

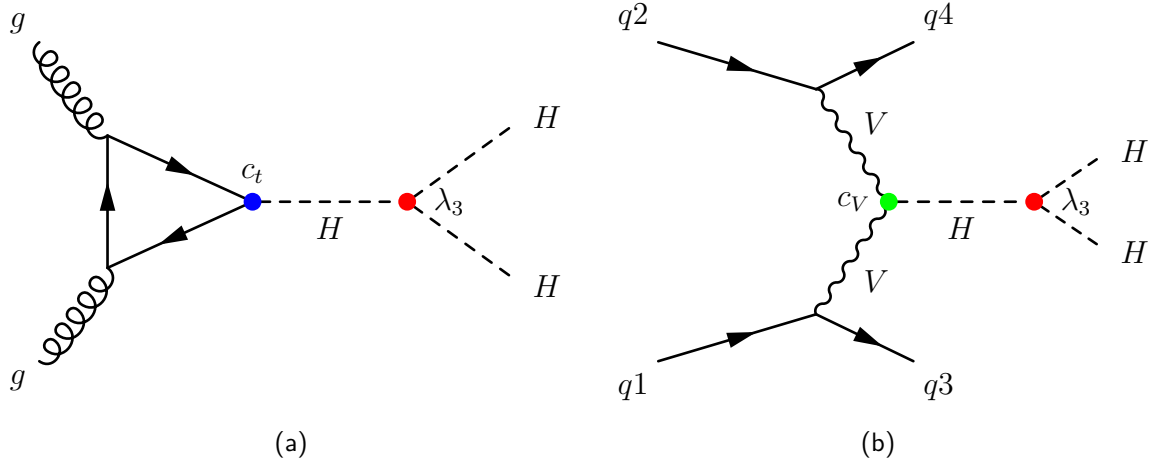


Figure 1.8: Feynman diagrams of the gluon-gluon fusion (ggF) (a) and vector boson fusion (VBF) (b) production modes of a pair of Higgs bosons that contains the trilinear Higgs self-coupling interaction vertex, characterised by the coupling strength λ_3 . The ggF production in (a) includes a vertex where the Higgs boson couples to a top quark labelled in blue, with coupling strength c_t . The VBF production in (b) relies on a vertex where the Higgs boson couples to a gauge boson V labelled in green, with a coupling strength of c_V .

With the growing dataset recorded at collider experiments as well as improvements in analysis techniques driven by machine learning advancements, the search for the production of triple Higgs bosons can be attempted to constrain λ_4 . The experimental constraints of its modifier κ_4 was carried out through the search for the production of three Higgs bosons using the ATLAS Run-2 dataset, forming a large proportion of this thesis in Chapter 4. The dominant SM production is ggF, with a cross-section of 0.079 fb at a center-of-mass energy of $\sqrt{s} = 13$ TeV and NNLO. Figure 1.9 shows the Feynman diagrams of all the leading-order ggF production modes as well as the interference between them. The diagrams are categorised in the pentagon and the box diagrams, labelled $\mathcal{P}$ and

$\mathcal{B}$ respectively, and the triangle diagrams labelled $\mathcal{T}_3$ and $\mathcal{T}_4$. Together, the production amplitude at leading order can be parametrised as

$$\mathcal{M} = \mathcal{P} + \kappa_3 \mathcal{B} + \kappa_3^2 \mathcal{T}_3 + \kappa_4 \mathcal{T}_4 \quad (1.30)$$

Detailed discussions on the interference terms can be found in [15]. The most significant contributions come from the box and pentagon diagrams, which are only affected by κ_3 . As for κ_4 , the associated terms largely contribute to the destructive interference, therefore reducing the sensitivity to it compared to κ_3 . However, large deviations from the $\kappa_3 = \kappa_4 = 1$ can change the destructive interference between the diagrams, therefore increasing the production cross-section predicted by the SM.

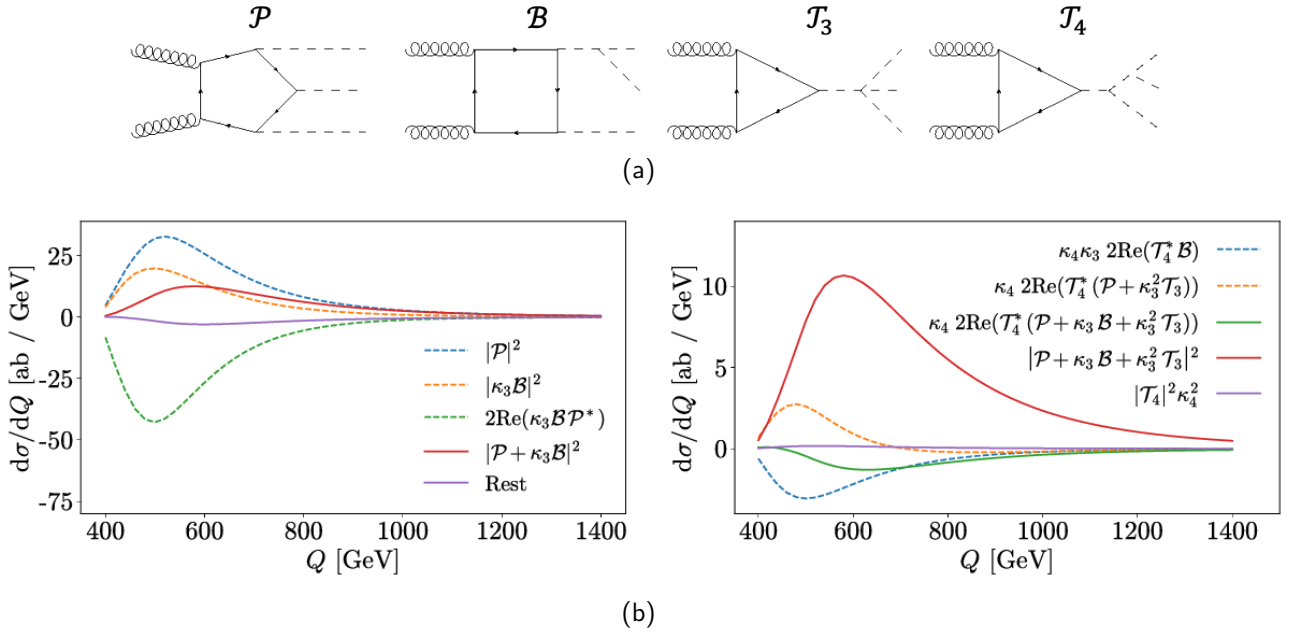


Figure 1.9: Feynman diagrams of leading-order tri-Higgs production in the ggF production mode (a). The contributions and interference between the different diagrams as a function of the invariant mass of the triple Higgs system Q is shown in (b). On the left, the contributions from the pentagon and box diagrams are shown in dashed lines, their destructive interference is shown in the green dashed line and the overall contribution in red solid line. On the right, the contributions from all diagrams are explicitly separated by the κ_4 dependence, and similar destructive interferences are shown in dashed lines. These figures are adapted from [15].

1.7 The Extended Higgs Sectors

Despite the success of the SM, there remain many open questions, such as the matter-antimatter asymmetry, neutrino oscillation, and dark matter. There are many BSM physics model that attempt to answer these questions from the theoretical perspective, predicting phenomenologies that could be measured from collider experiments. The extended Higgs sectors contain an array of BSM models that tackle the following questions:

1. **The hierarchy problem** The electroweak scale ($\sim 10^3$ GeV) and the Planck scale ($\sim 10^{18}$ GeV) where gravity becomes as strong as gauge couplings are separated by orders of magnitude. The mass of the Higgs boson is sensitive to quantum corrections from interactions with other particles via loop diagrams. These corrections scale quadratically with the ultraviolet energy cutoff Λ . However, if the SM is valid up to a very high scale and the observed mass of the Higgs boson is ~ 125 GeV, its bare mass must be a large fine-tuned number to cancel these corrections [16].
2. **The matter-antimatter asymmetry** The imbalance between the amount of matter and antimatter, or baryonic and antibaryonic matter, is unexplained. The measured amount of CP violation along with the deviation from thermal equilibrium during EWSB observed through cosmic background radiation measurements cannot account for this imbalance. Baryogenesis [17] is a popular theory that explains this large asymmetry, which favours a strong first-order EWPT that requires large deviations of the Higgs self-couplings from their SM values [6].
3. **Dark matter** The SM does not provide a viable candidate for dark matter that explains cosmological observations.
4. **Electroweak vacuum stability** There is ambiguity concerning the stability of the vacuum expectation value of the Higgs field. Current precise measurements of the Higgs boson properties and the top-quark mass suggest the existence of another minimum in the Higgs potential that is the global minimum, making the vev of the current Universe metastable. This is a consequence of the Higgs self-coupling decreasing monotonically and turning negative at high energy [18], indicating another decrease in gradient in the Higgs potential.

The extended Higgs sectors contain a wide array of BSM theories as potential answers to these questions. Many tackle them by extending the SM, hence the "extended" in the name. The most popular models include the Two Higgs Doublet Model (2HDM) which extends the one-doublet Higgs sector of the SM to contain at least two Higgs doublets. The $HHH \rightarrow 6b$ search presented later in this thesis is centred around a BSM model that extends the SM by two real singlets, namely the Two Real Singlet Model (TRSM). The above two models are outlined in some detail below.

The renormalisable scalar potential [19] of two Higgs doublets Φ_1 and Φ_2 in the 2HDM can be generalised to

$$\begin{aligned}
V(\Phi_1, \Phi_2) = & m_{11}^2 \Phi_1^\dagger \Phi_1 + m_{22}^2 \Phi_2^\dagger \Phi_2 - m_{12} (\Phi_1^\dagger \Phi_2 + \Phi_2^\dagger \Phi_1) \\
& + \frac{\lambda_1}{2} (\Phi_1^\dagger \Phi_1)^2 + \frac{\lambda_2}{2} (\Phi_2^\dagger \Phi_2)^2 \\
& + \lambda_3 \Phi_1^\dagger \Phi_1 \Phi_2^\dagger \Phi_2 + \lambda_4 \Phi_1^\dagger \Phi_2 \Phi_1^\dagger \Phi_2 + \frac{\lambda_5}{2} \left[(\Phi_1^\dagger \Phi_1)^2 + (\Phi_2^\dagger \Phi_2)^2 \right]
\end{aligned} \tag{1.31}$$

For two complex $SU(2)$ doublets, each contain 2 complex fields and 4 real degrees of freedom, resulting in a total of 8 fields. After EWSB, 3 fields give masses to the gauge bosons, with the remaining 5 fields as additional scalar particles arising from this extension. These scalar particles are: two CP-even scalars which include a SM-like Higgs boson (h) and a heavier scalar (H), one neutral CP-odd pseudoscalar (A) and two charged Higgs bosons ($H^\pm$). The main 2 types of 2HDM models are distinguished by the types of particles the two Higgs doublets couple to. In Type I, Φ_1 couples to no fermions and Φ_2 couples to all fermions. In Type II, Φ_1 couples to down-type quarks and leptons and Φ_2 couples to up-type quarks. Type II 2HDM is analogous to the Minimal Supersymmetric Model (MSSM) [20], which introduces a supersymmetric partner with opposite spin to all SM particles. It is motivated by the effective cancellation of the divergent contributions from quantum corrections to the mass of the Higgs boson from these supersymmetric partners, hence solving the hierarchy problem. The 2HDM models also offer possibilities for explicit or spontaneous CP violation [21] beyond the CKM mechanism, which could account for the observed baryon asymmetry. Other works have explored the origin of this asymmetry via baryogenesis in the 2HDM [22]. Loop corrections from the additional scalar particles it introduces as well as the complex mass terms mixing the two Higgs doublets can induce a first-order phase transition, therefore fuelling baryogenesis.

Other possible extensions to the SM includes singlet fields². The Two Real Singlet Model (TRSM) [24] extends the SM by two real singlet fields. Additionally, 2 $\mathbb{Z}_2$ symmetries are introduced to reduce the number of free scalar potential parameters. The result is three neutral CP-even scalar particles, X and S and the third identified as being the observed SM-like Higgs boson with a mass of 125 GeV [25]. The most general renormalisable potential for the singlets X and S and the SM-like Higgs doublet Φ can be expressed as

$$\begin{aligned}
V(\Phi, X, S) = & \mu_\Phi^2 \Phi^\dagger \Phi + \mu_S^2 S^2 + \mu_X^2 X^2 \\
& + \lambda_S S^4 + \lambda_X X^4 + \lambda_{XS} X^2 S^2 + \lambda_\Phi (\Phi^\dagger \Phi) \\
& + \lambda_{\Phi S} \Phi^\dagger \Phi S^2 + \lambda_{\Phi X} \Phi^\dagger \Phi X^2
\end{aligned} \tag{1.32}$$

where $\mu_\Phi, \mu_S, \mu_X, \lambda_\Phi, \lambda_S, \lambda_X, \lambda_{XS}, \lambda_{\Phi S}, \lambda_{\Phi X}$ are 9 real couplings. The additional scalar particles are viable candidates for dark matter. Within the allowed regions of parameter space, a strong first-order phase transition is allowed, catalysing a successful baryogenesis which again could account for the observed baryon asymmetry in the Universe. At the LHC, the three neutral scalar particles from this extension can be probed through their decay products. The decays can occur via a symmetric topology, where two identical scalar particles are produced, an asymmetric decay, where different scalar particles are produced, or a cascade decay where the heaviest scalar particle decays to the lighter ones successively. Experimentally, these decays lead to final states that contain three or four SM-like Higgs bosons, where the former is targeted by a search performed at the ATLAS experiment, detailed in Chapter 4.

The extended Higgs sectors enrich the space of collider phenomenology by introducing a variety of BSM models that predict new scalar particles. They provide a plethora of solutions to some of the open questions about the fundamentals of the Universe. Experimentally, many of them occur at a rate that is within reach of the current collision energy at the LHC, which fuels an extensive search programme for these processes.

²An interesting study [23] performing a global fit on gravitational wave signals found that extended singlet models can simultaneously induce a strong first-order phase transition and satisfy the observed dark matter abundance.

2 | Experiment

Trust me, I'm literally an ATLAS author.

— Harry Simpson

The majority of the work presented in this thesis, namely Chapters 3 and 4, is based on data collected at the ATLAS experiment, one of the four experiments situated in the CERN accelerator complex that houses the Large Hadron Collider (LHC). Specifically, the proton-proton collision data collected during the period of 2015-2018, corresponding to the LHC Run-2, is used for these studies. A general introduction of the LHC is provided in Section 2.1. Then the details of the ATLAS detector and its sub-detectors are presented in Section 2.2. The Run-2 and Run-3 configurations of the trigger system, crucial to the ATLAS experiment data-taking, are described in detail to give context for related studies mentioned later in this thesis. Finally, the reconstruction of various physical objects from electric signals collected in the detector is outlined in Section 2.5.

2.1 The Large Hadron Collider

Situated on the French-Swiss border close to Geneva, the Large Hadron Collider is the largest particle accelerator and collider on the planet, built by the European Organization for Nuclear Research (CERN). Its circumference spans 26.7 km, with an average depth of approximately 100 m underground. Two oppositely circulating beams of protons are brought to collision at a centre-of-mass energy reaching a maximum of 13.6 TeV during the most recent data-taking period. The highly energetic beams are bent by superconducting magnets that generate a strong magnetic field. The protons used in Run-2 data taking are sourced from hydrogen atoms with their electrons stripped away in an electric field in Linac2. They are then accelerated by radio frequency (RF) cavities, and in stages using synchrotrons. They operate by changing the direction of the electromagnetic field which regulates the acceleration of the protons according to their position in the field. The result is bunches of protons focused by these cavities and accelerated to higher energies according to their frequency. The protons are first accelerated to 50 MeV before reaching the Proton Synchrotron Booster (PSB), then to 1.4 GeV before reaching the Proton Synchrotron (PS) which further accelerates them to

25 GeV. Finally the Super Proton Synchrotron (SPS) pushes the energy to 450 GeV before injecting the beams into the beam pipes of the LHC. The injected beams consist of bunches of 10^{11} protons separated by 25 ns. The individual hard-scattering interactions are referred to as *events* as the two proton beams collide. Circulating in the beam pipe, the protons beams are accelerated to their maximum energy, then focused and brought to collide at four collision points around the LHC ring. At these collision points situate the four major experiments at CERN: ALICE, ATLAS, CMS and LHCb. The focus of this thesis is the ATLAS detector, a general purpose detector that targets a wide variety of physics signatures. Together with various other accelerator components and smaller collider experiments, they constitute the CERN accelerator complex, shown in Figure 2.1.

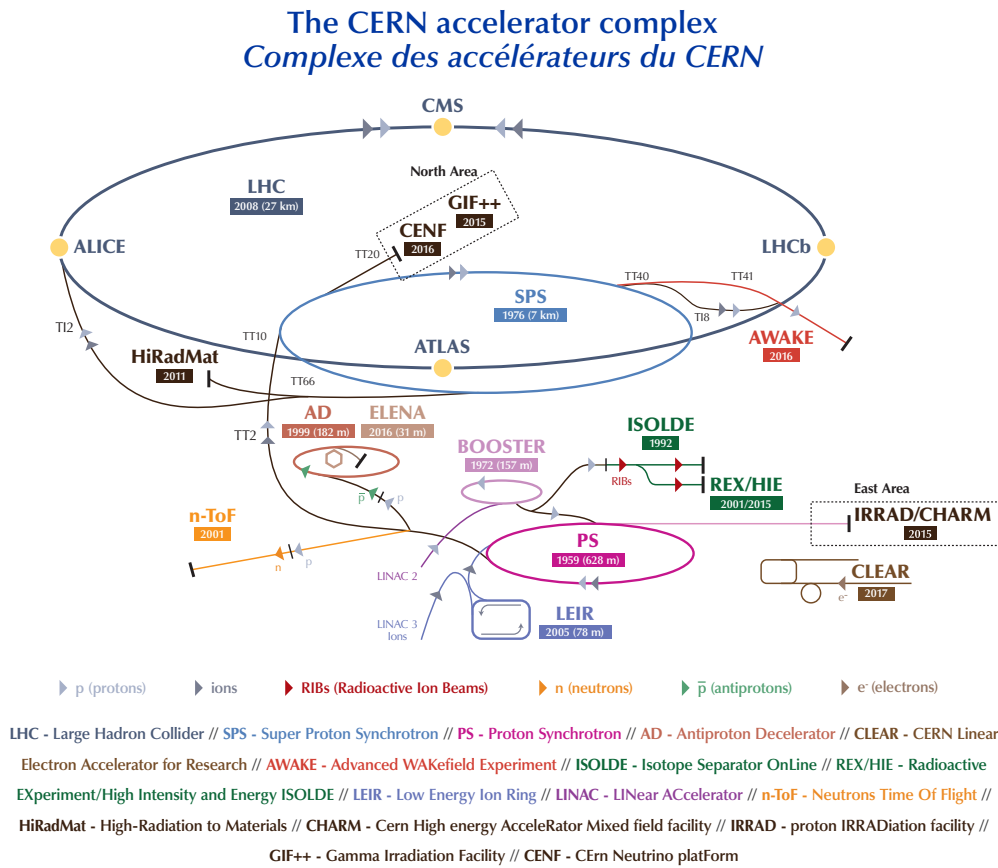


Figure 2.1: A schematic of the CERN accelerator complex, showing the main LHC ring in black, around which the four major experiments situated at the collision points, ALICE, ATLAS, CMS and LHCb, are highlighted by yellow circles. The proton sources, different stages of the synchrotron acceleration and the other smaller experiments are also shown. The figure is taken from [26].

2.2 The ATLAS Detector

The ATLAS (A Toroidal LHC Apparatus) detector is a general purpose detector that aims to detect a wide variety of physics signatures produced at the LHC. The sub-detectors of this apparatus are responsible for detecting different physics objects. They are the inner detector, which records the trajectories, or tracks, of charged particles; the electromagnetic calorimeter (ECAL) which detects particles that interact via the electromagnetic force such as electrons and photons; the hadronic calorimeters (HCAL) which measures particles that interact hadronically via the strong force such as pions and kaons, and the muon spectrometer which detects muons as they are much heavier than the electrons and therefore are not stopped inside the ECAL. Invisible particles such as neutrinos are not directly measured. Instead, they contribute to the missing momentum in the transverse plane which is the negative sum over the momenta of all observed particles, which would be zero if all particles are accounted for in the centre-of-mass frame. A schematic of the ATLAS detector and its sub-detectors is shown in Figure 2.2.

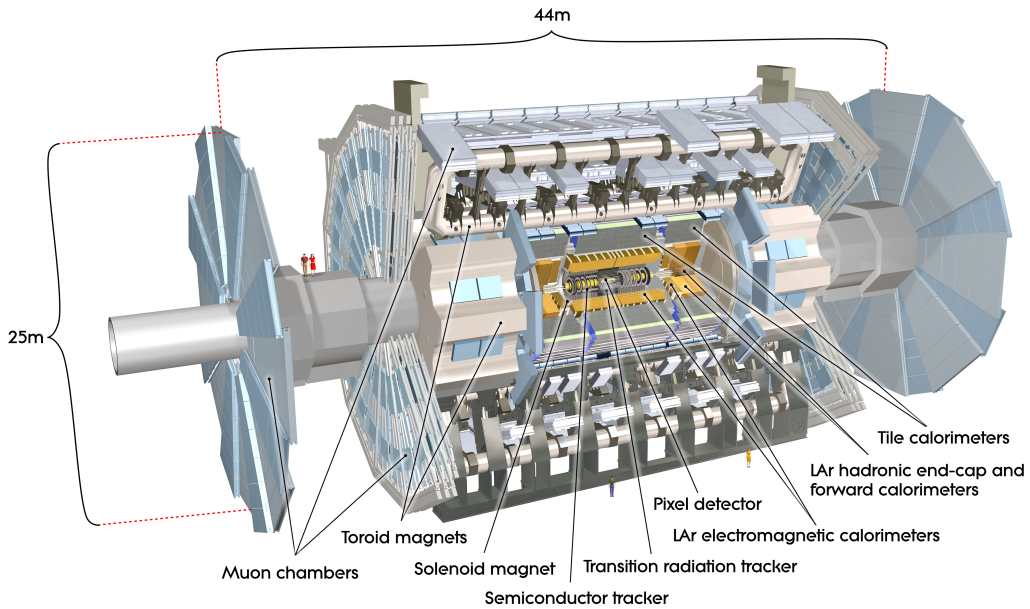


Figure 2.2: A schematic of the ATLAS detector and its various sub-detectors. The figure is taken from [27].

A word on the coordinate system used by the ATLAS experiment for ease of reference in later chapters. The system adopted is a right-handed coordinate system shown in Figure 2.3, where the

origin is the proton-proton interaction point and the z -axis runs along the beam axis. The *transverse* plane is the $x - y$ plane perpendicular to the beam axis, commonly used to summarise the kinematic features of physical objects and collision events. This plane is used as it discounts any contribution in the longitudinal direction that is dependent on the momenta of the incoming particles which is often not known. The polar angle θ is defined with respect to the z -axis, where $\theta = 0$ is the direction along the z -axis and $\theta = 90$ is perpendicular to it. The azimuthal angle ϕ is defined with respect to the x -axis, where $\phi = 0$ is the direction along the x -axis and $\phi = 90$ is perpendicular to it. An important variable, the *pseudorapidity* $\eta = \ln \tan \left(\frac{\theta}{2} \right)$ is often used instead of θ , where $\eta = 0$ is the direction perpendicular to the beam axis, and $\eta \rightarrow +\infty$ is the direction along the beam axis. The angular separation between two objects ΔR can be defined by combining the difference in their pseudorapidity and azimuthal angles, that is $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$. It is also often used to describe the angular size of a jet cone.

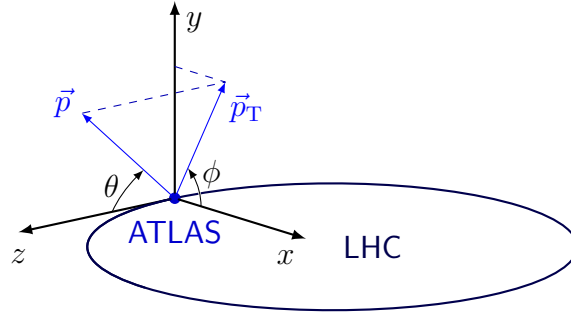


Figure 2.3: Schematic of the right-handed coordinate system used by the ATLAS experiment. The z -axis is aligned with the beam axis, and the polar angle θ and the azimuthal angle ϕ are defined with respect to the z and x axes respectively. An example of a momentum vector $\vec{p}$ and its corresponding transverse momentum $\vec{p}_T$ is also illustrated. The diagram is adapted from [28].

2.2.1 The Inner Detector

As suggested by its name, the inner detector is the inner-most part of the ATLAS detector. Its proximity to the collision point makes it suitable for obtaining precise measurements of the tracks left by electrically charged particles or their subsequent decays. The measurement of these tracks is referred to as tracking. In order to measure their masses and momenta, it is immersed in a 2 T magnetic field, provided by a solenoid magnet. The inner detector covers a region of $|\eta| < 2.5$ with a pixel detector, a semiconductor tracker (SCT), and a transition radiation tracker (TRT),

illustrated in Figure 2.4. The pixel and SCT sub-detectors provide high-precision measurements of the particle tracks that enable the reconstruction of decay vertices and other important variables, while the TRT records a large number of tracking points per particle at lower resolution.

The pixel detector consists of silicon pixel modules installed in three layers in a cylindrical barrel, and three layers in the endcaps. Also as a part of the pixel detector, an Insertable B-Layer (IBL) [29] was installed during the shutdown of the LHC in 2013 - 2014, 33 mm away from the beam axis. The main motivation for its installation was to maintain tracking performance ahead of LHC Run-2 at higher instantaneous and integrated luminosities and foreseeing accumulated radiation damage to the silicon detector. It has important implications for particle identification that relies heavily on tracking and vertex reconstruction downstream such as flavour-tagging, discussed in Chapter 3 and used to conduct the physics analysis in Chapter 4. The addition of high-precision hits closer to the interaction point significantly improves the extrapolation resolution of tracks compared to the pixel detector in Run-1 [29]. Hence the resolution of the impact parameters in the transverse (d_0) and longitudinal (z_0) directions, both used as inputs to flavour-tagging algorithms, are improved.

The SCT [30] consists of 4,088 silicon strip modules, arranged in four layers in a cylindrical barrel, and in nine disks in each endcap. In each module, the two pairs of silicon strips are attached at a 40 mrad angle to provide 2 measurements which are combined to construct space-points. To construct a particle originating in the beam-interaction region, typically 4 space-point measurements are provided. The modules are arranged in 12 rings in the barrel, and in 3 rings in the endcaps. The strips in the barrel are uniformly separated by 80 μm , measured between the centres of adjacent strips. The strips in the endcaps run radially with an angular separation of 161.5 μrad .

The TRT [31] is a tracking drift chamber in the outermost layer of the inner detector, consisting of 52,544 tubular drift tubes (or straws). It provides tracking information from the straw tubes, as well as electron identification via the transition radiation from fibres between the straw tubes. The tubes are filled with a gas mixture of 70% X_e , 27% CO_2 and 3% O_2 . Each with a diameter of 4 mm, they are arranged in three layers in a cylindrical barrel around the beam axis. The barrel

covers a region corresponding to $|\eta| < 1$.

The resolution of the three sub-detectors in the $r - \phi$ and the z directions are summarised in Table 2.1.

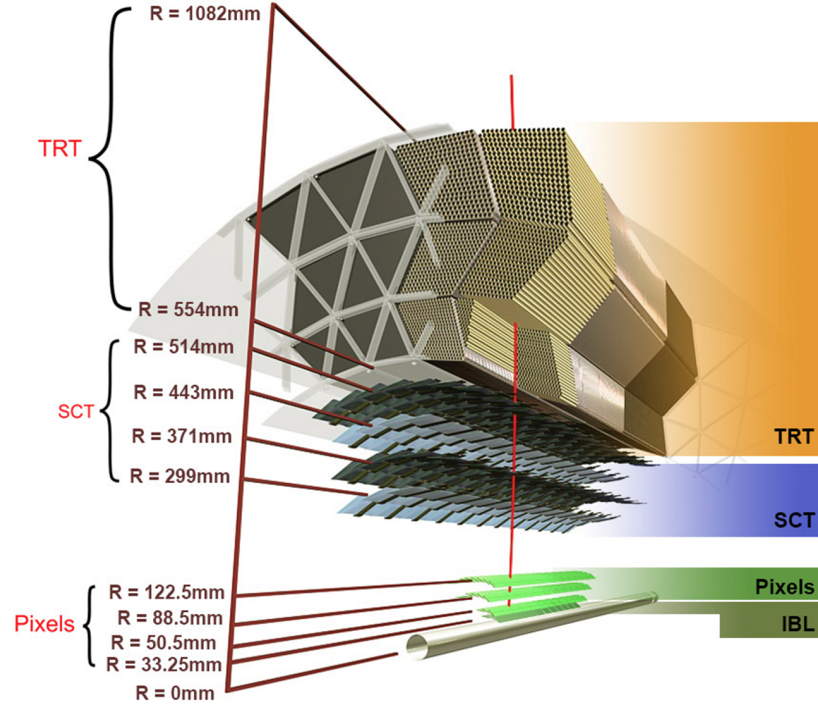


Figure 2.4: A cross-section of the ATLAS inner detector, its components and their dimensions. The figure is taken from [32].

Sub-detectors	Resolution in $r - \phi$ [μm]	Resolution in z [μm]
IBL	10	60
Pixel	10	115
SCT	17	
TRT	130	

Table 2.1: Table summarising the intrinsic resolutions in the $r - \phi$ and the z directions of the sub-detector components of the ATLAS inner detector. They are taken from [32]. The resolution in z for the SCT and TRT is coarse and not relevant here as they are further away from the beam axis.

2.2.2 The Electromagnetic Calorimeter (ECAL)

The ECAL is responsible for measuring the precise energy of electrons and photons, covering a pseudorapidity range of $|\eta| < 3.2$. It consists of passive material which forces the particles to lose their energy due to bremsstrahlung or pair production, and active material which measures the deposited energy as a signal. Lead absorption plates are used as the passive material in the ECAL. The passive material tends to have a large radiation length X_0 , which is the mean distance over which an electron or a photon travels in the material before depositing an energy of $\frac{E}{e}$, ensuring efficient electromagnetic showering and good resolutions. The active materials convert the energy loss into a signal either via ionisation or scintillation. Lead-liquid Argon is used for this purpose for its linear energy response, forming the Liquid-Argon Layers (LAr). In the barrel region, together with the readout electronics, they are arranged in an accordion-like geometry that naturally provides a full coverage in the ϕ direction [33]. The particles traversing through the active material ionise the liquid argon. The amount of ionisation is linearly proportional to the energy deposited by the particles, and creates an electronic signal to be read out.

The ECAL has a dimension of approximately $22 X_0$ in the barrel region, and $24 X_0$ in the endcaps, which is large enough to capture full electromagnetic showers. The barrel covers a pseudo-rapidity range of $|\eta| < 1.475$, and the endcaps cover $1.375 < |\eta| < 3.2$.

2.2.3 The Hadronic Calorimeter (HCAL)

The HCAL captures the hadronic decays of particles that interact via the strong force. Located behind the ECAL, the HCAL barrel covers a pseudo-rapidity region of $0.8 < |\eta| < 1.7$, while the endcaps cover $1.5 < |\eta| < 3.2$. The barrel is more finely divided into the central barrel and extended barrel regions. It consists of steel as the passive material, and scintillating tiles as the active material. The tiles are arranged radially and normal to the beam axis [33]. The radiation depth of the tiles are approximately 7.4 interaction lengths, defined as the mean distance a high-energy hadron travels through a material before undergoing hadronic showering. In contrast to electrons and photons, the nuclear radiation lengths are much longer than the radiation lengths. Therefore the central HCAL

is 5.8 m in length. The light signals produced by hadronic showers in the tiles are then read out by photomultiplier tubes. In the endcaps, copper is used as the passive material and LAr is used as the active material. In the most forward region where $3.1 < |\eta| < 4.9$, the passive material changes to tungsten that provides more precise energy measurements.

2.2.4 The Muon Spectrometer

The muon spectrometer is located at the outermost layer of the ATLAS detector, covering a pseudorapidity range of $|\eta| < 2.7$. It is dedicated to measuring hits left by traversing muons, which are substantially heavier than electrons, and therefore are not stopped in the ECAL. Again, in order to measure the tracks of charged particles, it is immersed in a magnetic field. In the range of $|\eta| < 1.4$, the magnetic field is provided by the large barrel toroid magnet, and for $1.6 < |\eta| < 2.7$, it is provided by two smaller endcap magnets. In the barrel region, muon tracks are measured in chambers arranged in three cylindrical layers surrounding the beam. In the transition and endcap regions, they are measured in planar chambers also arranged in three layers perpendicular to the beam. The spectrometer has excellent resolution for the momentum of charged particles, of roughly 10% for 1 TeV tracks [33], and charge identification. Monitored Drift Tubes (MDTs) are largely responsible for this high momentum resolution with small tube diameters and constituting a large proportion of the spectrometer in the barrel region ($|\eta| < 2.7$). In the forward region of $2 < |\eta| < 2.7$, Cathode Strip Chambers (CSCs) are arranged closest to the beam pipe to provide high granularity and good timing resolution while withstanding radiation damage. Triggering using muons is provided by Resistive Plate Chambers (RPCs) in the barrel region, and the Thin Gap Chambers (TGCs) in the endcaps. They contain gas between the plates that gets ionised by the traversing muons, initiating an avalanching of charged particles that are subsequently captured by the plates. The resolution is roughly 1 cm by 1 ns, which provides information on the muon tracks at a high rate for the hardware trigger logic to recognise their multiplicity and energy [33].

For ATLAS Run-3, the New Small Wheels (NSWs) [34] were installed in the forward muon tracking region. It addresses the increased muon trigger rate and degraded tracking performance at an increased instantaneous luminosity. It consists of 8 layers of Micro Patterns (micromegas) and

small-strip TGCs to provide tracking and triggering capabilities. The new installation provides fast precision tracking at a rate of 15 kHz cm^{-2} , and a spatial resolution of less than $100 \text{ }\mu\text{m}$ per detection plane.

2.3 Energy, Luminosity and pile-up

The centre-of-mass energy is given by $\sqrt{s}$ where s is the total energy and momentum of the two initial state particles [35], defined as

$$s = \left(\sum_{i=1}^2 E_i \right)^2 - \left(\sum_{i=1}^2 \mathbf{p}_i \right)^2 \quad (2.1)$$

The rate at which a given physics process occurs, often referred to as the *cross-section*, is dependent on s . Another important parameter of the collider is the integrated *luminosity* L . It can be used to calculate the number of events expected for a given process with cross-section σ_{event}

$$N = L\sigma_{event} \quad (2.2)$$

The instantaneous luminosity $\mathcal{L}$ [36], [37] is integrated luminosity differentiated with respect to time. It can also be fully determined by parameters of the beam, that is

$$\begin{aligned} \mathcal{L} &= \frac{N_1 N_2 n_b f}{4\pi\sigma_x\sigma_y} \\ &= n_b \frac{\langle\mu\rangle f}{\sigma_{inel}} \end{aligned} \quad (2.3)$$

which describes the instantaneous luminosity of two Gaussian-distributed beams, where N_1, N_2 are the number of colliding particles per bunch from the two beams, n_b is the number of bunches in each beam, f is the revolution frequency, which is 11246 Hz for proton-proton collisions, and σ_x and σ_y are the standard deviations of the two Gaussian beams in the transverse direction. The instantaneous luminosity can also be expressed in terms of the pile-up parameter $\langle\mu\rangle$, used to quantify the average number of inelastic collisions per bunch pair. The parameter σ_{inel} is the cross-section of the inelastic collisions. At the LHC, the two beams collide at a small angle which reduces the amount of collisions

by a geometric reduction factor F . The total integrated luminosity recorded at the ATLAS detector during LHC Run-2 data-taking amounted to 140.01 fb^{-1} [37], at a centre-of-mass energy of 13 TeV. For the Run-3 data-taking as of the end of 2024, the total recorded luminosity reached 183 fb^{-1} at a centre-of-mass-energy of 13.6 TeV. It aims to record up to 300 fb^{-1} of collision data, significantly surpassing Run-2. The cumulative integrated luminosity recorded at the ATLAS detector for proton-proton collisions so far in LHC Run-3 data-taking as of the end of year 2024 is shown in Figure 2.5. The average pile-up recorded for Run-3 as of the end of 2024 at the ATLAS detector is shown and compared with the average for Run-2 in Figure 2.6. The average number of interactions per bunch crossing increased significantly in Run-3, attributing to the large increase in luminosity.

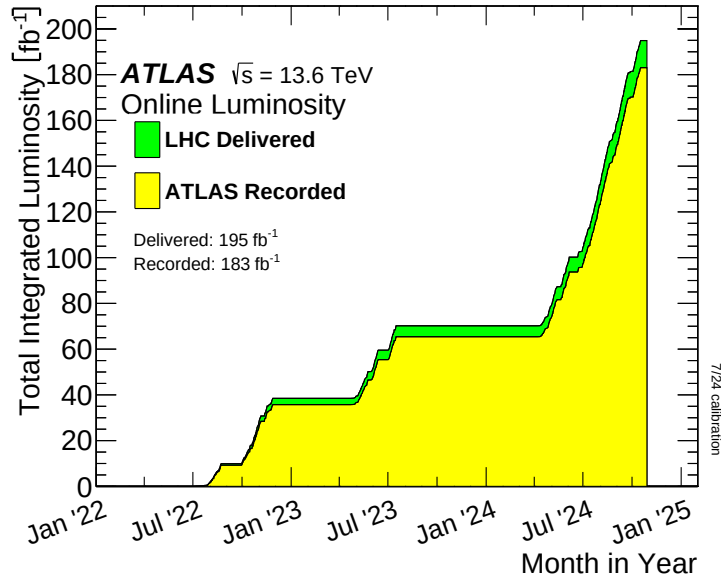


Figure 2.5: Cumulative luminosity delivered by the LHC (green) and recorded at the ATLAS detector (yellow) during stable beams of proton-proton collisions at a centre-of-mass energy of 13.6 TeV during the LHC Run-3 data-taking period between 2022-2024. The figure is taken from [38]

2.4 The ATLAS Trigger System

In the ATLAS detector, bunches of protons are brought to collision at an extraordinary rate of 40 million times per second. In practice, it is not possible to reconstruct and record every collision event to disk, therefore a trigger system is put in place to only select events that are interesting for various physics analyses to be stored. Whether an event is considered interesting depends on the particles produced from the collision. They leave characteristic signatures in the detector that are revealed

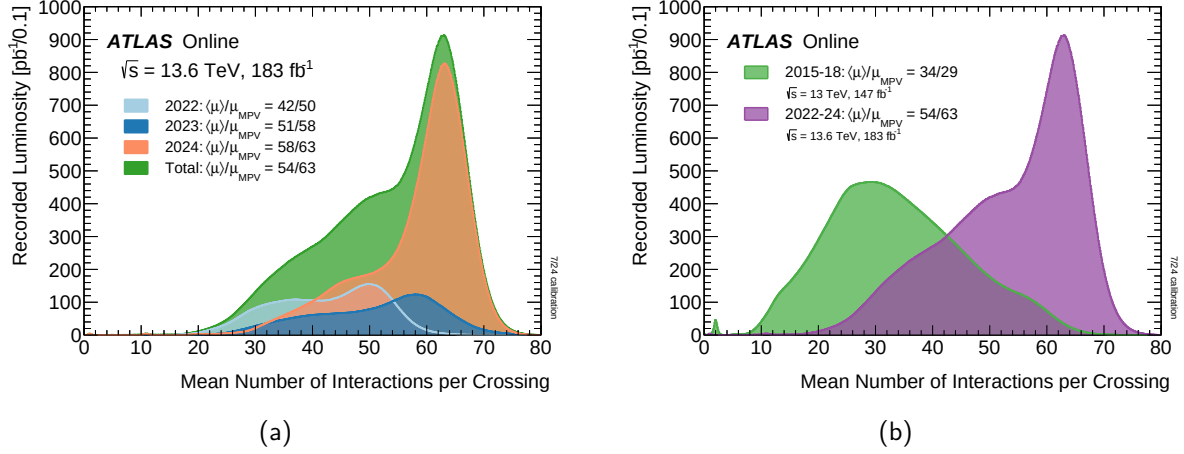


Figure 2.6: The luminosity-weighted distributions of the average pile-up parameter $\langle\mu\rangle$, defined as the mean number of interactions per bunch crossing, measured at the ATLAS detector during proton-proton collisions for the LHC Run-3 data-taking period between 2022-2024 (a), and compared to the LHC Run-2 data-taking period between 2015-2018 (b). The figures are taken from [38].

after fast reconstruction at the trigger level (*online* reconstruction) using information collected from various parts of the detectors described in Section 2.2. A final number of approximately 1000 events per second are selected to be written to disk and prepped for more detailed reconstruction (*offline* reconstruction) further downstream. The triggering process in the ATLAS detector during Run-2 data-taking is performed in two stages:

- **Level 1 Trigger (L1)** - a hardware-level trigger which utilises information from the calorimeter and muon spectrometer to reduce the number of events to 100,000 per second.
- **High Level Trigger (HLT)** - a software-level trigger, which performs more detailed reconstructions of particles using various dedicated algorithms and reduces the number of events to 1000 per second to be recorded to disk.

The following sections give an overview of these two systems. Details on the performance of specifically the Run-2 trigger configuration can be found in [39].

2.4.1 Level-1 Trigger (L1)

The L1 trigger relies on information collected in the calorimeter and the muon spectrometer. The L1 calorimeter (L1Calo) trigger [40] identifies electrons, photons, τ -leptons and jet candidates above various thresholds, as well as global sums of missing transverse energy. Trigger towers of dimension

$(\Delta\eta \times \Delta\phi) = (0.1 \times 0.1)$ collected from cells in the ECAL and HCAL are used as inputs. The L1 muon (L1Muon) trigger uses hits in the muon spectrometer to identify muon candidates. The overall L1 trigger selects events based on their event-level quantities, multiplicity of objects above given thresholds, and their topological features [39]. A decision is formed by combining information from L1Calo and L1Muon triggers [41], as well as other detector subsystems. The maximum number of events selected per second is limited to 100,000, and the decision for an event is made in 2.5 μ s. In addition to making a per-event trigger decision, L1 also identifies Regions-of-Interest (Rols) in η and ϕ within the detector, which are 3-dimensional regions of the detector. The information from various sub-detectors of the accepted events at L1 can be accessed during the next trigger stage in the HLT when requested. Only selected events and Rols found by the L1 trigger are processed in the HLT to minimise computing time and cost.

2.4.2 High Level Trigger (HLT)

The second level of the ATLAS trigger system is a software-based trigger, the High Level Trigger (HLT) [39]. It constitutes a mixture of fast algorithms providing early rejection of events to reduce the CPU consumption and time for more precise reconstruction algorithms that follow. To execute all the algorithms in the HLT, a computing farm consisting of 40,000 processing units is used. The algorithms are designed to extract the relevant features of specific types of signatures from within the Rols identified by the L1 trigger. For certain signatures, such as b -jets, information from the full detector is required. Trigger decisions are made on a per-event basis based on whether all conditions for a given trigger are satisfied. Events that satisfy all trigger conditions are read out at an average rate of 1.2 kHz during ATLAS Run-2 data-taking. They are then stored in permanent storage for offline reconstruction at the CERN computing centre. The trigger conditions that need to be satisfied at both the L1 and HLT levels are summarised in trigger chains, specific to the signatures that are targeted by different physics analyses. In each data-taking year, the trigger chains used are designed according to the targeted physics programme, which constitute a trigger menu. Changes in trigger menu are necessary during the run in response to any changes in the data-taking conditions, algorithms and physics interests.

2.5 Physical Object Reconstruction

The electronic signals generated by different particles recorded in the detector are used to reconstruct physical objects such that their properties as well as the overall collision events can be analysed. The reconstruction of these objects occurs at the trigger level as well as the offline level. This section will focus on the offline object and event reconstruction.

2.5.1 Tracks, Vertices and Calorimeter Topo-clusters

Charged particles are detected in the inner detector by leaving helical trajectories in the inner detector due to the presence of a magnetic field. The accurate reconstruction of these trajectories is important for providing precise measurements of their momentum, charge information and decay vertices. Clusters of energy deposited by traversing charged particles can be identified and collected as *hits* [42] from the pixel detector and the SCT. They are defined as a group of pixels or strips in a sensor with energy deposit yielding a charge above a certain threshold [43]. The charge is often collected from multiple adjacent pixel sensors, where the number increases with pseudorapidity due to the detector geometry. The spatial separation between charged particles spans only a few pixels. A track seed of a charged particle is formed by joining a set of three hits [43]. A combinatorial Kalman filter [44] is used to form track candidates by combining the track seeds with other compatible hits from remaining layers of the pixel and SCT detectors. This creates multiple track candidates per seed, resulting in high tracking efficiency at a relatively low CPU cost. However, the large volume of charged particles traversing through the detector makes it an extremely dense environment which leads to ambiguity in the association between hits and tracks due to overlaps. To resolve this ambiguity, a track score is calculated for each track candidate, which takes into consideration the track quality, intrinsic resolutions of the sub-detectors, track momentum and multiplicity [42]. The track candidates are then sorted by their scores, and iteratively, ones failing to meet certain criteria of a minimum p_T threshold, $|\eta| < 2.5$, minimum hits in pixel and SCT detectors and minimum impact parameters¹ are rejected [42]. These criteria ensure a low rate of tracks reconstructed from

¹The transverse (d_0) and longitudinal (z_0) impact parameters are defined as the distances of closest approach from a track to the measured beam spot position, in the transverse and longitudinal directions.

a random combination of hits, often referred to as fake tracks. In addition, a neural network [43] is trained to identify merged clusters (energy clusters created from multiple particles) with an efficiency of 85%. The final track candidates are then defined using a global χ^2 re-fitting procedure that estimates track parameters at high precisions. The tracks are then extrapolated in the TRT and re-fit over the hits to further improve momentum resolution [42]. For particles produced at a greater distance from the beam line, a secondary track reconstruction (back-tracking) is performed using the remaining detector hits. The procedure starts with hits in the TRT compatible with Rols seeded by energy deposits in the ECAL. They are then extended to consider hits in the pixel and SCT detectors and track candidates are formed by again using the Kalman filter, and the hit-track assignment ambiguity solved using steps described above.

After track candidates are identified and associated with hits, the underlying interaction point, the *primary vertex*, is reconstructed. The method consists of two steps, vertex finding and vertex fitting. In Run-2, an iterative multi-vertex finding algorithm was used to estimate the initial position of the vertex in the z direction before performing a fit that minimises the χ^2 between compatible tracks and the vertex position iteratively. It was replaced by an adaptive multi-vertex fitter algorithm [45] to handle the higher density of vertices due to the higher pile-up conditions in Run-3. The initial vertex position of a track is instead determined using a Gaussian resolution model to increasing accuracy. The primary vertex of an event is then defined as the one with the largest sum of squares of transverse momenta of the associated tracks [46].

Signals left in the calorimeters from electromagnetic and hadronic showers are collected as three-dimensional topological cell clusters of the calorimeters, also known as calorimeter topo-clusters [47]. A clustering algorithm groups these cells and removes adjacent ones with insufficient signal. The information in these clusters is associated the energy and location of the showers, and is used to correct the reconstruction of physics objects such as jets and missing transverse energy. These corrections include the effects of calorimeter response to hadrons, energy loss in inactive material between calorimeter modules and signal loss outside of the formed clusters [47]. They are also crucial for seeding jet reconstruction both online and offline, as well as studying the jet sub-structures.

2.5.2 Jets

A large amount of quarks are produced at the LHC from the collision of protons. They do not exist in unbound states, and therefore quickly hadronise and shower in the detector, creating sprays of particles called *jets*. *b*-jets, which are jets containing *b*-quarks, are of particular importance as the *b*-quark is the heaviest quark that can be observed in isolation. It is therefore often the final product of the decay of massive exotic particles. The dominant decay mode of the Higgs boson is also to a $b\bar{b}$ pair, making it a statistically advantageous mode to search for.

Jet constituents - Particle Flow objects To reconstruct a jet, first its constituents such as topo-clusters, tracks and vertices are collected from the full detector, and combined into a global object, the Particle Flow object (PFlow) [48]. An algorithm forms these objects using the following steps:

1. Tracks meeting a set of criteria specified in [48] are selected, ordered by descending p_T .
2. Each track is matched to a single topo-cluster.
3. The expected energy deposited by the particle is calculated based on the location of the topo-cluster, and the momentum of the associated track. First the algorithm attempts to match each track to a single topo-cluster. Then it evaluates the probability of multiple topo-clusters being associated with a track, and adds topo-clusters accordingly to estimate the total energy.
4. Finally, the energy that contributed to the track is subtracted from the topo-clusters, forming a PFlow object.

The procedure is validated using simulated single-pion and dijet MC samples. Combining information from the tracks and calorimeter topo-clusters has several advantages. Tracks are only left by charged particles, whereas neutral particles deposit their energy in the calorimeter. Tracks have high precision in particle momentum and position measurements at low energy, whereas calorimeter topo-clusters have higher energy resolution at high energy.

Jet clustering - the anti- k_t algorithm PFlow objects are then clustered in to jets (PFlow jets) using a jet clustering algorithm, the anti- k_t algorithm [49]. It clusters jets into a conical topology that is infrared and collinear safe, where infrared safety ensures additional soft gluon radiation does not change jet structure and boundaries, and collinear safety means the splitting of a particle into two collinear ones does not change the jet structure either. The algorithm works by first defining a distance metric d_{ij} between a particle i and a pseudojet j , and a distance metric between the same particle and the beam B :

$$\begin{aligned} d_{ij} &= \min(k_{t,i}^{2p}, k_{t,j}^{2p}) \frac{\Delta_{ij}^2}{R^2} \\ d_{iB} &= k_{t,i}^{2p} \end{aligned} \tag{2.4}$$

where k_t is the transverse momentum of a particle or a jet. The quantity Δ_{ij} is a geometrical scale, $\Delta_{ij} = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where y and ϕ are the rapidity and the azimuthal angles. The variable p is the relative power of the energy with respect to Δ_{ij} , and R is the radius parameter of the jet cone. All pair-wise d_{ij} distances are calculated and compared to d_{iB} . If $d_{ij} < d_{iB}$, the particle i is clustered in the pseudojet j , forming a new pseudojet. If $d_{ij} > d_{iB}$, the particle i is removed from the list of particles to be clustered. The power of the anti- k_t algorithm lies in its choice of the power parameter p . For $p = 1$, this is equivalent to the inclusive k_t algorithm [50], where the particles with low k_t (soft particles) are clustered first. For $p = 0$, the inclusive Cambridge/Aachen algorithm [51] is recovered, where the clustering is purely dependent on the angular separation between the particle and the pseudojet. The anti- k_t algorithm uses $p = -1$, which essentially merges particles with high k_t (hard particles) first. This approach is beneficial for a number of reasons. The distance d_{ij} when $p = 0$ becomes large when the soft particles are in close proximity, leading to clustering of soft particles with hard ones before the clustering among soft particles. As a result, a hard particle that does not have a neighbouring hard particle within a radius of $2R$ would accumulate all soft particles in a perfect jet cone. If such a neighbouring hard particle exists, $p = 0$ cannot result in two conical jets due to the overlap between them. Therefore setting $p = -1$, equivalent to prioritising the clustering of the hard particles, ensures the jet shape is ultimately affected by the hard particles instead of the soft ones, hence making the jet boundary resilient against soft radiations.

Types of jets Jets defined in the ATLAS experiment are categorised by their radius parameter R . The most common three types are small-R jets, where $R = 0.4$, large-R jets where $R = 1.0$, and variable radius track (VR) jets where R is adjusted depending on the jet p_T .

Small-R jets are commonly used in analyses that target well-separated, or resolved, jets. For example, the $HHH \rightarrow 6b$ analysis in this thesis searches for a final state that contains 6 small-R jets. They are required to pass a Jet Vertex Tagger (JVT) [52] criterium, a tagger based on the k -nearest neighbour (kNN) algorithm, which outputs the probability of a jet arising from hard-scatter or pile-up. Further jet quality criteria are imposed to minimise background of jets reconstructed from energy deposited by beam induced background, cosmic rays and calorimeter noise.

Large-R jets are often used in analyses that search for objects with p_T higher than its rest mass, also referred to as boosted objects. As they decay hadronically, their decay products become collimated, making it difficult to separately resolve them. Instead, they are reconstructed in a single jet cone with a larger radius parameter R , such as in the boosted $HH \rightarrow 4b$ analyses. Jet trimming [53], a procedure that removes soft radiation, is often used to mitigate contamination from initial state radiation, multiple interactions and pile-up.

VR jets [54] have varying R defined by $R = \frac{\rho}{p_T}$, where ρ is a constant that determines the rate of change in R . This reflects the nature of the showering of hadrons from a parton leading to a cone of fixed angular size regardless of its orientation in the detector, which is not captured by a fixed radius parameter. This approach is also motivated by the improvement in signal efficiency when used in the boosted regime. They are also often used to study the substructure of large-R jets and used in b -tagging in the boosted regime.

Jet energy correction The energy of the reconstructed PFlow jets needs to be corrected due to effects such as pile-up and detector interactions. Two quantities are measured, the jet energy scale (JES), defined as the ratio between the measured energy and the true energy $JES = \frac{E_{meas}}{E_{true}}$, and the jet energy resolution (JER), defined as the ratio between the standard deviation of the measured energy fitted as a Gaussian distribution and the true energy $JER = \frac{\sigma(E_{meas})}{E_{true}}$. To measure the JES, the jet momentum is measured from the jet recoiling against a reference object such as a photon

or a Z boson whose momentum can be measured precisely, or a reference jet in a dijet system where the momenta are well-balanced. In Run-2, the jet energy scale and resolution are measured in simulated events and compared between PFlow jets and truth jets. Truth jets are reconstructed using only stable particles in the final state, while excluding muons, neutrinos, and particles arising from pile-up. The jet energy, four-momentum, p_T , and mass are then corrected to match the value at the particle level using a calibration. For data recorded in Run-2, the jet energy scale calibration follows the following steps [55]:

1. Pile-up corrections are applied to remove excess energy due to pile-up, consisting of a correction based on the p_T density of the event and its jet area, and a correction derived from MC simulation in the residual that is a function of the average interactions per bunch-crossing ($\langle\mu\rangle$) and the number of reconstructed primary vertices in the event (N_{PV}).
2. An absolute MC-based calibration is applied in energy and direction such that they agree with the truth jet in simulated dijet MC events.
3. Global sequential calibration [56] (GSC) removes the dependence of global jet observables such as information related to tracking and the muon chambers, and energy deposits in the calorimeters. This improves the jet p_T resolution without changing the average jet energy response.

For data recorded in Run-3, the GSC is replaced with a neural network based calibration method (GNNC) [57]. It was found that the kinematic variables used by GSC are uncorrelated, therefore limiting the set of corrections that can be applied. A deep neural network is therefore trained to include additional observables at the jet and event levels while accounting for their correlations, resulting in a more accurate calibration. Furthermore, a separate network is trained in each $|\eta^{det}|$ region to account for the difference in the calorimeter geometry across it. The closure in the jet p_T from this calibration is typically within 1% and provides a smoother correction than the previous GSC method.

Jet Reconstruction in the Trigger The reconstruction of jets follows a slightly different procedure in the trigger. At the L1 trigger, jets are identified from trigger tower inputs from the ECAL

and HCAL within the central part of the detector ($|\eta| < 2.5$), and at a coarser granularity for $|\eta| > 2.5$ [40]. These trigger tower clusters also contain the coordinates of the Rols that are selected and passed onto the HLT.

Jets in the HLT are reconstructed using a clustering algorithm called the anti- k_t algorithm [49] as particles contained within a cone. The algorithm takes the full set of calorimeter cell information as inputs, and the energy of the clustered jets is then corrected by subtracting the contributions from pile-up events. Calibration factors that are dependent on E_T and η derived from simulations are used for Run-2, and an in-situ calibration procedure is used for Run-3 data-taking [58]. These calibration steps correct for the pile-up radiation, the flavour-dependence of jet features and the jet 4-momentum. For proton-proton collision events, the default cone radius for jets is 0.4 (small-R jets), although jets with radius of 1.0 are also used (large-R jets) by analyses that consider boosted signatures where the jets are more collimated. The reconstructed jets are required to pass certain E_T thresholds in L1 trigger and the HLT separately. Jets that contain b -hadrons are called b -jets, and they are of particular interest to many physics analyses and are a crucial element of several studies covered in this thesis. The identification and selection of them in the HLT is described in detail in Section 3.1.

b -jets Jets containing b -hadrons, which are the products of the hadronisation of b -quarks, are referred to as b -jets. Many interesting physics processes contain b -quarks in the final state, as they are the heaviest quarks that are observed in isolation. The heaviest quark, the top-quark, has an extremely short lifetime and undergoes weak decay before forming any bound states as mesons or baryons. The b -quark however can hadronise to form b -hadrons, which typically have a lifetime on the orders of picoseconds. This relatively long lifetime allows them to travel a measurable distance in the detector before decaying into lighter quarks. The identification of the resulting b -jets is called *b -tagging*, and dedicated machine learning algorithms are developed for this task. These b -tagging algorithms exploit the unique characteristics of the b -hadron decay, illustrated in Figure 2.7. They include the presence of a displaced secondary vertex as it decays to lighter quarks, high decay track multiplicity, and large transverse impact parameters of the decay tracks. The algorithms typically

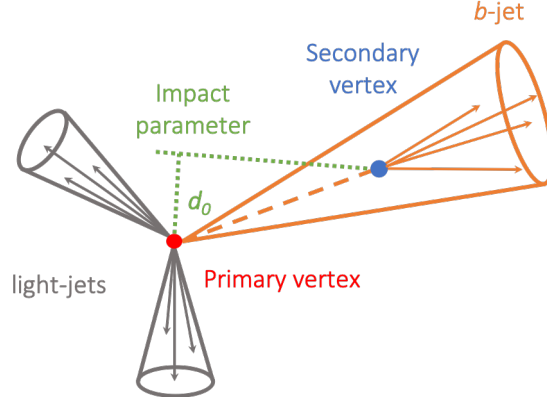


Figure 2.7: Schematic of a b -hadron decay and some of its characteristics exploited by b -tagging algorithms.

perform classification between three jet flavours: b -jets, c -jets and light-flavour jets. The outputs of b -tagging algorithms are jet flavour probabilities p_b, p_c, p_{light} , which can be intuitively interpreted as the probability of a jet being one of the three jet flavours. A discriminant score is calculated by combining them into a log-likelihood ratio

$$D_b = \ln \left(\frac{p_b}{f_c * p_c + (1 - f_c) * p_{light}} \right) \quad (2.5)$$

where f_c is a tunable parameter related to the effective c -jet fraction in the sample used to train the algorithm. A jet is defined as a b -jet if its discriminant value is above a certain threshold. These thresholds are commonly referred to as working points, and are used to define b -tagging efficiency ε_b , which is the fraction of b -jets correctly identified in simulated $t\bar{t}$ and Z' MC samples. The algorithms rely heavily on the accurate reconstruction of tracks and vertices. The latest iteration of the b -tagging algorithm is based on a Graph Neural Network and transformers architecture which utilises jets and their associated tracks as inputs. The evolution and details of the b -tagging algorithms are discussed later in Section 3.2.

b -jet calibration As b -tagging algorithms are trained on simulated MC events, the differences in the b -tagging performance between MC and data are corrected using a dedicated calibration procedure. The correction applied to MC is derived as a *scale factor* (SF) for each jet flavour, defined as the ratio of the b -jet efficiencies between data and MC $SF = \frac{\varepsilon_f^{data}}{\varepsilon_f^{MC}}$ where f is the jet flavour tagged by the algorithm, namely b -jet, c -jet and light-flavour jet. A separate set of SFs is

derived for each jet flavour [59]–[61]. The calibration for the b -jet efficiency of b -jets uses simulated $t\bar{t}$ events which are enriched in b -jets from the top-quark decay, and derives SFs in discrete bins of jet p_T as well as working points. A new technique [62] based on optimal transport is also developed for data recorded in Run-3, which allows for a continuous calibration across p_T and working points. This facilitates the derivation of corrections to higher dimensions using transport maps. The calibration for the b -jet efficiency of c -jets, which is just the fraction of c -jets mis-tagged as b -jets, is derived using semi-leptonic $t\bar{t}$ events where one of the W bosons decays into a charm-quark with a sufficiently large branching fraction. The calibration for the b -jet efficiency of light-flavour jets uses simulated multijet MC events. The main assumption in the measurement of the mis-tagging rate here is that the signed impact parameter of tracks associated with a light-flavour jet is approximately symmetric around zero. Therefore the mis-tagging rate of light-flavour jets can be obtained by evaluating the tagging algorithm after flipping the impact parameter sign of the tracks.

2.5.3 Muons

Muons are used in the work presented in this thesis as the semi-leptonic decay of a b -quark decay to a muon leaves a characteristic signal in the detector, where a muon can be cleanly reconstructed within the jet cone of a b -jet. This is useful for selecting data events enriched in b -jets for calibration purposes. Muons are reconstructed primarily using tracks in the muon spectrometer that are matched to tracks in the inner detector [63]. Tracks in the muon spectrometer are reconstructed by identifying local track segments using a Hough transform, before combining them into track candidates. The candidates are made three-dimensional by adding information from precision measurements in the bending plane and the second coordinate from the trigger detectors. A global χ^2 fit is then performed on the muon trajectory through the magnetic field, along with an iterative ambiguity solving and re-fitting step. The resulting tracks from the muon spectrometer are then matched to tracks identified in the inner detector to perform a combined fit based on hits in the two detector systems. Information in the calorimeters is also used to account for energy loss as well as to determine the track parameters. Several working points are defined by balancing the muon identification efficiency, background rejection and muon track resolution.

3 | Rejecting $g \rightarrow b\bar{b}$ Splitting in the ATLAS b -jet Trigger

What a lot of people don't realise is that
gunshots are louder than techno.

— Dan Guest

3.1 The ATLAS b -jet Trigger

Bottom quarks (b -quarks) are common in the final states of many interesting physics processes, particularly in those containing the Higgs boson, as its most common decay products are a pair of b - and anti- b -quarks. The ATLAS b -jet triggers use b -tagging algorithms to select events containing b -jets as part of the HLT for various physics analyses targeting b -quarks in the final state. The identification of these events requires the precise reconstruction of the primary vertex (PV finding) and tracks (tracking) in each event at the trigger level. The steps taken to form a b -tagging decision in the HLT are as follows:

- **Fast tracking:** tracks are reconstructed for all jets with $E_T < 20$ GeV within composite Regions of Interest (SuperRols) that combine Rols in regions of $|\eta| < 0.4$, $|\phi| < 0.4$ found in the hardware trigger (L1).
- **Fast b -tagging:** a b -tagging algorithm is applied to the reconstructed tracks in order to select events for more precise tracking, a CPU intensive step. Events with no b -tagged jets are rejected from further tracking.
- **Full scan tracking and PV finding:** tracks from the entire Inner Detector (ID) are combined with topological clusters in the calorimeter to form Particle Flow (PFlow) jets. The primary vertex of each event is formed by taking the vertex with the largest sum over the squared p_T of the assigned tracks.
- **Precision tracking:** PFlow jets with E_T over a certain threshold depending on the specific trigger are selected for further processing. More precise track reconstruction is performed for each PFlow jet in an Rol centred on the jet direction and the z -position of its primary vertex.

Crucial components for b -tagging such as the transverse and longitudinal impact parameters as well as secondary vertices are reconstructed at this stage.

- **Precision b -tagging:** the reconstructed quantities and tracks are used as inputs to a final b -tagging algorithm, which outputs a b -tagging score for each jet in the event.

A b -jet trigger chain specifies the b -tagging algorithm used at the HLT level and the working point associated with it which informs the b -tagging score threshold.

3.2 b -tagging Algorithms

The ATLAS b -tagging algorithms have undergone several iterations of development, mostly driven by advancements in machine learning algorithms. Starting as a Boosted Decision Tree (BDT), namely MV2 [64] in Run-2, it evolved into several Deep Neural Networks (DNNs) DL1r and DL1d [64], and now a Graph Neural Network (GNN) GN1 [65] in Run-3. The previous two architectures relied on a set of low-level algorithms that constructed discriminating variables from features unique to the b -hadron decay using the tracks and jets formed by the decay products. The outputs of these low-level algorithms are then combined to form the inputs to a high-level algorithm, either a BDT or a DNN, which then performs jet flavour classification. The low-level algorithms are grouped into two categories: impact-parameter-based and secondary-vertex-based algorithms. Impact-parameter-based algorithms include IP2D/IP3D, which use the transverse and longitudinal impact parameter significances of tracks to output log-likelihood ratios of the b -jet, c -jet and light-jet flavour hypotheses. To account for correlations between decay tracks, a Recurrent Neural Network (RNNIP) and later a Deep Sets Neural Network [66] (DIPS [67]) is trained which only uses track-level information to output a b -jet hypothesis. Secondary-vertex-based algorithms include SV1, which reconstructs a single displaced secondary vertex in each jet by iterating over all possible two-track vertices and rejecting ones compatible with the decay of long-lived particles, photon conversions or hadronic interactions with the detector. It outputs eight discriminating variables associated with the reconstructed vertex. Another algorithm JetFitter [68] reconstructs the entire decay topology of the b -hadron inside a

jet including the secondary and tertiary vertices to eliminate effects of tracks from pile-up events ¹. A high-level algorithm such as a DNN takes the outputs of these low-level algorithms as inputs and performs classification between b -jets, c -jets and light-flavour jets, where b -jets are considered as the signal flavour, c -jets and light-flavour jets are considered background flavours. More recently, a GNN removes the dependency on these low-level algorithms completely, and directly performs classification using information provided only by the tracks and jets of the decay products. The outputs are jet-flavour probabilities p_b , p_c and p_{light} , they are combined into a single b -jet discriminant in the form of a log-likelihood ratio in Equation 2.5. Jets with scores above a chosen threshold of D_b are tagged as b -jets, and the fraction of genuine b -jets is defined as the b -jet efficiency ε_b , analogous to the true positive rate (TPR). The performance of the algorithm is characterised by its ability to reject c -jets and light-flavour jets at a given ε_b . Several fixed b -jet efficiencies are chosen as working points (WPs) where the background rejections are evaluated, defined as the inverse of the fraction of c -jets or light-flavour jets being mis-identified as b -jets above a given threshold of D_b , analogous to the inverse of the false positive rate (FPR). The b -jet discriminant D_b of GN1 deployed in the ATLAS b -jet trigger are shown in Figure 3.1. Their c -jet and light-flavour jet rejections are often shown as functions of b -jet efficiency, commonly referred to as a Receiver Operating Characteristic (ROC) curve, such as in Figure 3.2 for GN1.

3.3 $g \rightarrow b\bar{b}$ Rejection Algorithm

The b -jet class in the ATLAS b -tagging algorithms includes jets containing one (single- b -jets) or two b -hadrons (bb -jets). However, these two types of jets can arise from different physics processes. Single- b -jets can result from the decay of a Higgs boson, where two b -quarks are produced with sufficient angular separation such that two individual single- b -jets are reconstructed. Whereas bb -jets are often produced from small-angle gluon splitting ($g \rightarrow b\bar{b}$) where the gluon is sufficiently energetic, common in proton-proton collision events at the LHC. The latter is usually not targeted by physics analyses, and thus regarded as background jets. Rejecting them would improve signal

¹Pile-up events are events in addition to the collision events of interest in the detector. They can be additional proton-proton collisions (in-time pile-up), collisions just before or after bunch-crossing of proton beams (out-of-time pile-up), or collisions between protons and detector material or gas particles.

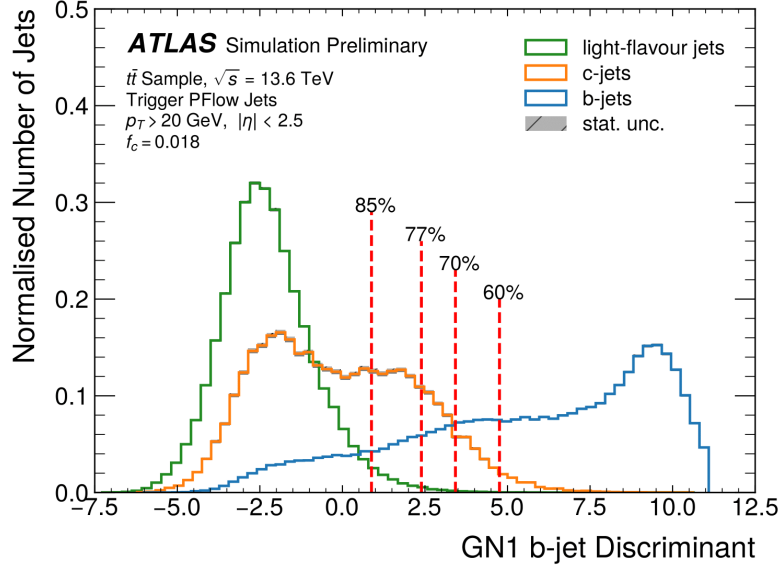


Figure 3.1: Distributions of the b -jet discriminant for the GN1 algorithm, of b -jets, c -jets and light-flavour jets in simulated $t\bar{t}$ events. All jets are HLT Particle Flow jets, and are required to have $p_T > 20$ GeV and $|\eta| < 2.5$. The red dashed vertical lines represents working points at four chosen b -jet efficiencies. The figure is taken from [3].

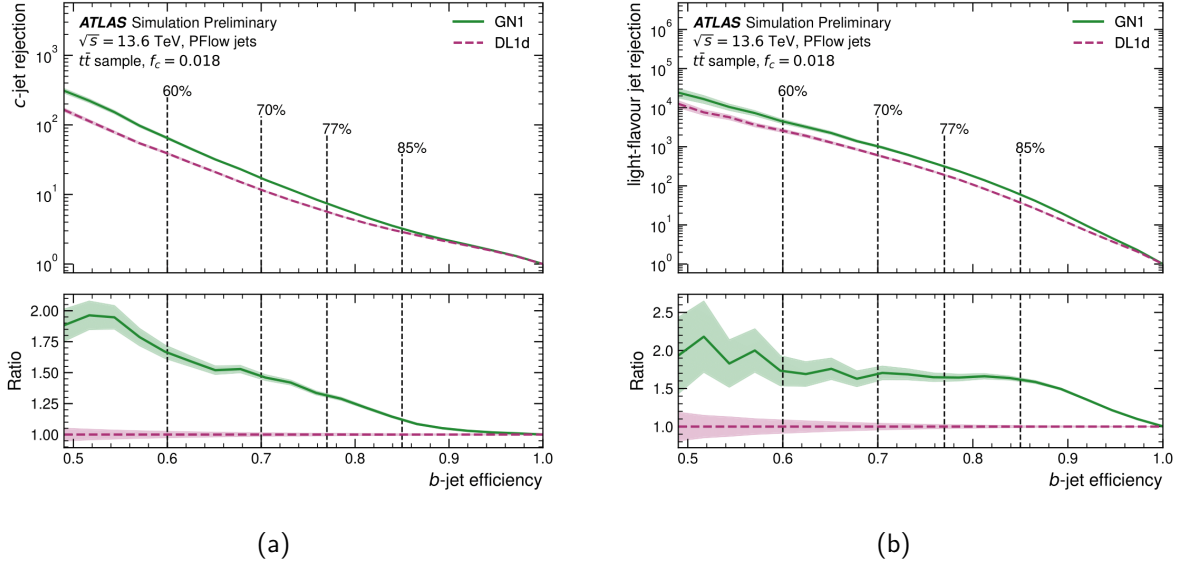


Figure 3.2: Rejection of c -jets (a) and light-flavour jets (b) as functions of b -jet efficiency of the GN1 algorithm (green solid) compared to the benchmark DL1d algorithm (purple dashed). The algorithms are evaluated on HLT Particle Flow jets in simulated $t\bar{t}$ events. The dashed vertical lines represents working points at four chosen b -jet efficiencies. The figures are taken from [3].

purity for analyses targeting single- b -jets. It is also important to reject them at the trigger level, as a reduction in the readout rate would be valuable for the trigger bandwidth. The identification of bb -jets as signal by the b -tagging algorithm in the b -jet trigger is evident in the b -jet discriminant distribution of GN1 in Figure 3.3, where a large proportion of distribution of bb -jets overlaps with the single- b -jets. As a result, most bb -jets present in data would be identified as b -jets, and the events containing them are more likely to pass the b -tagging requirement in the trigger and be kept.

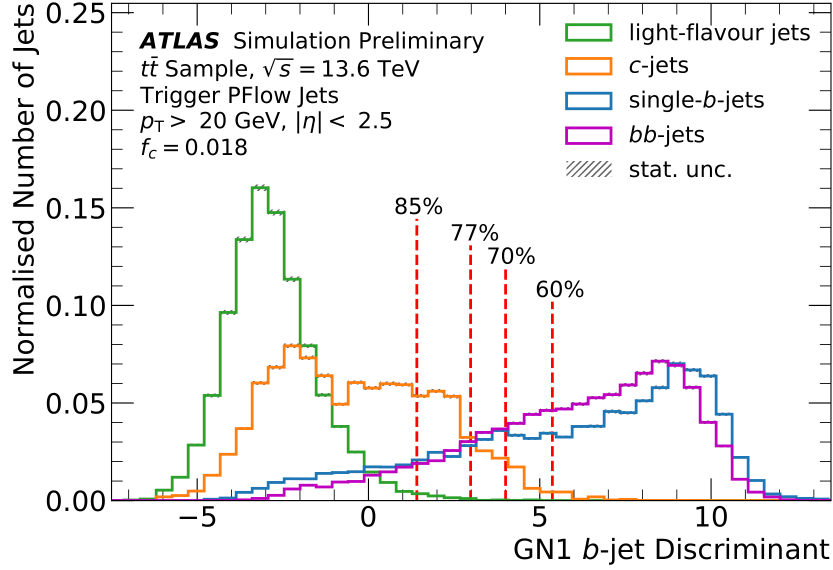


Figure 3.3: Distributions of the b -jet discriminant for the GN1 algorithm, of single- b -jets, bb -jets, c -jets and light-flavour jets in simulated $t\bar{t}$ events. All jets are HLT Particle Flow jets, and are required to have $p_T > 20$ GeV and $|\eta| < 2.5$. The red dashed vertical lines represents working points at four chosen b -jet efficiencies. The figure is published in [3].

As a solution, a neural-network-based algorithm, DL1dbb, is trained to further separate jets that are identified as b -jets by the current b -tagging algorithm into single- b -jets and bb -jets. The dedicated algorithm to bb -jet rejection is motivated by the fact that it is more challenging for the algorithm to separate these two jet flavours while maintaining a good level of c -jet and light-flavour jet rejection. The DL1dbb algorithm is designed to be used in conjunction with the baseline b -tagging algorithm in the trigger. The outputs of the algorithm are the jet flavour probabilities for single- b -jets and bb -jets, which are combined into a single discriminant in the form of a log likelihood ratio:

$$D_b^{DL1dbb} = \ln \left(\frac{p_b}{p_{bb}} \right) \quad (3.1)$$

Trigger jets passing the baseline b -tagging algorithm, namely GN1, that also pass DL1dbb would thus

have a higher purity of single- b -jets.

3.3.1 Training Setup

The algorithm is trained on $t\bar{t}$ and QCD multijet samples, simulated in proton-proton collision at a centre-of-mass energy of $\sqrt{13}$ TeV. The $t\bar{t}$ sample is simulated using POWHEG interfaced with PYTHIA8. The QCD multijet sample is simulated using POWHEG [69] interfaced with PYTHIA8 [70] and using EVTGEN package [71] to handle particle decays. These samples are selected as their events are enriched in both single- b -jets and bb -jets covering a wide range of p_T . They are also common processes produced in proton-proton collisions at the LHC. Jets in these samples are required to have $p_T > 20$ GeV and $|\eta| < 2.5$ to ensure they are reconstructable and within the tracking acceptance region of the detector. To ensure balanced jet flavour composition and equal distributions in p_T and $|\eta|$, jets in the training samples are resampled. This prevents the trained model to perform better on the jet flavour or in a particular p_T or $|\eta|$ region with larger statistics. The statistics in bb -jets are limited compared to single- b -jets, which are five times more abundant in these samples. In order to balance the statistics between the two jet flavours while maximising the bb -jet statistics in the training samples, the single- b -jets were resampled to match the shape of the p_T and $|\eta|$ distributions of bb -jets. The ratio between the two jet flavours in each bin of the two-dimensional distribution in p_T and $|\eta|$ is used as a sampling probability in that bin. Once the shapes of the distributions between the two jet flavours match, the bb -jets are then up-sampled to match the statistics of the single- b -jets. The jets in the training and validation samples are orthogonal, but selected and sampled in the same way. The training sample consists of 5 million jets, and the validation sample consists of 1 million jets in total.

The algorithm is a fully-connected feedforward deep neural network [72]. The training is carried out using the ATLAS flavour-tagging software Umami [73] using a Keras backend. Single- b -jets are labelled as the signal class, and bb -jets as the background class. Binary classification of the two classes is performed using supervised training, where the signal and background classes are assigned truth labels of 1 and 0 respectively. The neural network consists of 2 hidden layers containing 12 and 6 hidden nodes in each, which consist of weights and biases as trainable parameters. During each iteration of the training dataset, two jet flavour probabilities p_b and p_{bb} are provided as outputs

for each event. A binary cross-entropy loss function is used to compute the difference between the outputs and their target values y_b and y_{bb} :

$$L = -[y_b \log(p_b) + y_{bb} \log(1 - p_{bb})] \quad (3.2)$$

where $y_b + y_{bb} = p_b + p_{bb} = 1$.

The trainable parameters of the neural network are updated to minimise this loss after each iteration of the training dataset. The details of the neural network architecture is summarised in Table 3.1.

Hyperparameters	Values
Number of input variables	44
Number of hidden layers	2
Number of nodes per layer	[12, 6]
Number of training epochs	300
Learning rate	0.001
Training mini-batch size	15000

Table 3.1: Hyperparameters used in the training of the DL1dbb algorithm

3.3.2 Input Variables

The input variables of the DL1dbb algorithm are chosen to capture the kinematic differences between single- b -jets and bb -jets. As bb -jets contain two b -hadrons instead of one, two secondary vertices are present within the jet cone at points where the b -hadrons decay. As a result, kinematic variables reconstructed at the secondary vertices are important in separating these two jet flavours. bb -jets also contain more tracks within the jet cone as they have more particles from the decay of the two b -hadrons. The same set of input variables used in the training of the baseline b -tagging algorithm in the b -jet trigger at the time DL1d were found to be sufficiently discriminating, and are summarised in Tables 3.2 and 3.3. The kinematic differences between single- b -jets and bb -jets are well-captured by these variables, and their distributions can be found in Appendix A. Since bb -jets contain two b -hadrons in a single jet cone, it has a higher number of charged tracks and therefore a higher reconstructed invariant mass at the two secondary vertices, captured by the $SV1_{masssvx}$ variable

shown in Figure 3.4(a). They also tend to have a lower fraction of energy contribution from the tracks at the secondary vertices as additional tracks are likely to be produced from initial state radiation from the gluons. This is captured by the bb -jets having smaller values of the variable $JFSV_{efrac}$, shown in Figure 3.4(b). However, the differences between single- b -jets and bb -jets are smaller than between single- b -jets, c -jets and light-flavour jets, making this classification task more challenging. It motivated the strategy of separating this classification from the baseline b -tagging task of DL1d in order to avoid compromising the rejection of c -jets and light-flavour jets, which are more abundant in the data compared to bb -jets.

Input	Variables	Description
Kinematics	p_T	Jet p_T
	$ \eta $	Absolute jet rapidity
DIPS	DIPS _{pb} DIPS _{pc} DIPS _{pu}	DIPS probability output of a given jet being a b -jet, c -jet and light-flavour jet
SV1	SV1 _{masssvx}	Invariant mass of tracks at the secondary vertex (SV)
	SV1 _{efracsvx}	Energy fraction of the tracks associated with the SV
	SV1 _{NGTinsvx}	Number of tracks used in the SV
	SV1 _{N2Tpair}	Number of two-track vertex candidates
	SV1 _{Lxy}	Transverse distance between the PV and SV
	SV1 _{L3d}	Distance between the PV and the SV
	SV1 _{significance3d}	Distance between the PV and the SV divided by its uncertainty
	SV1 _{deltaR}	ΔR between the jet axis and the direction of the SV relative to the PV

Table 3.2: Input variables from DIPS and SV1 and their definitions used in the training of the DL1dbb algorithm.

Input	Variables	Description
JetFitter	JF_{mass}	Invariant mass of tracks from displaced vertices
	JF_{efrac}	Energy fraction of the tracks associated with the displaced vertices
	JF_{nVTX}	Number of vertices
	$JF_{nTracksAtVtx}$	Number of tracks from multi-prong displaced vertices
	$JF_{nSingleTracks}$	Number of single-prong displaced vertices
	$JF_{N2Tpair}$	Number of two-track vertex candidates
	$JF_{significance3d}$	Significance of the average distance between PV and displaced vertices
	JF_{deltaR}	ΔR between the jet axis and the vectorial sum of momenta of all tracks attached to displaced vertices
JetFitter Secondary Vertices	$JFSV_{mass}$	Invariant mass of tracks at the secondary or tertiary vertex
	$JFSV_{energy}$	Energy of the tracks associated with the secondary or tertiary vertex
	$JFSV_{efrac}$	Energy fraction of the tracks associated with the secondary or tertiary vertex
	$JFSV_{nTracks}$	Number of tracks associated with secondary or tertiary vertex
	$JFSV_{displacement2d}$	Transverse displacement of the secondary or tertiary vertex
	$JFSV_{displacement3d}$	Distance of secondary or tertiary vertex from the PV
	$JFSV_{\eta_{trk}^{min}, \eta_{trk}^{max}, \eta_{trk}^{avg}}$	Minimum, maximum, and average relative rapidity of tracks at secondary or tertiary vertex
	$JFSV_{\eta_{alltrk}^{min}, \eta_{alltrk}^{max}, \eta_{alltrk}^{avg}}$	Minimum, maximum, and average relative rapidity of all tracks

Table 3.3: Input variables from JetFitter and secondary vertices reconstructed from JetFitter, and their definitions used in the training of the DL1dbb algorithm.

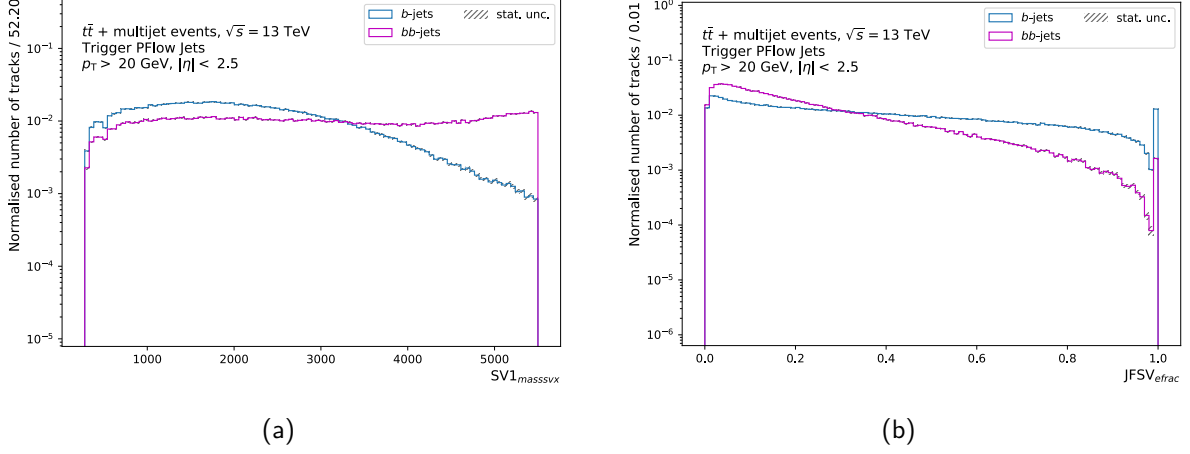


Figure 3.4: Two input variables used in the training of DL1dbb, $SV1_{masssvx}$ and $JFSV_{efrac}$, which are representative of the kinematic differences between single- b -jets and bb -jets.

3.3.3 Performance

The training loss and accuracy over the 300 training epochs are shown in Figure 3.5, which demonstrates good training convergence. The performance of the trained model is evaluated on the validation dataset, and the distribution of its discriminant D_b^{DL1dbb} is shown in Figure 3.6 for single- b -jets and bb -jets. At the 77% b -jet efficiency working point, a good level of separation between the two jet flavours is achieved. When used in conjunction with GN1 at the fixed 85% b -jet efficiency working point, DL1dbb is able to achieve a factor of 2 in bb -jet rejection at an overall b -jet efficiency of 77%, as shown in Figure 3.7. The overall b -jet efficiency is calculated by multiplying the b -jet efficiencies at the chosen working points of the two algorithms.

3.4 Working Point Optimisation

Since the DL1dbb algorithm rejects a different background jet flavour from the baseline b -tagging algorithm, the working points of the two algorithms need to be tuned to ensure a good level of rejection across all background jet flavours. Since the overall b -jet efficiency is calculated by multiplying the b -jet efficiencies at the chosen working points of the two algorithms, the working point of the baseline algorithm needs to be increased in order to make way for the addition of the DL1dbb algorithm to maintain the same overall efficiency as before.

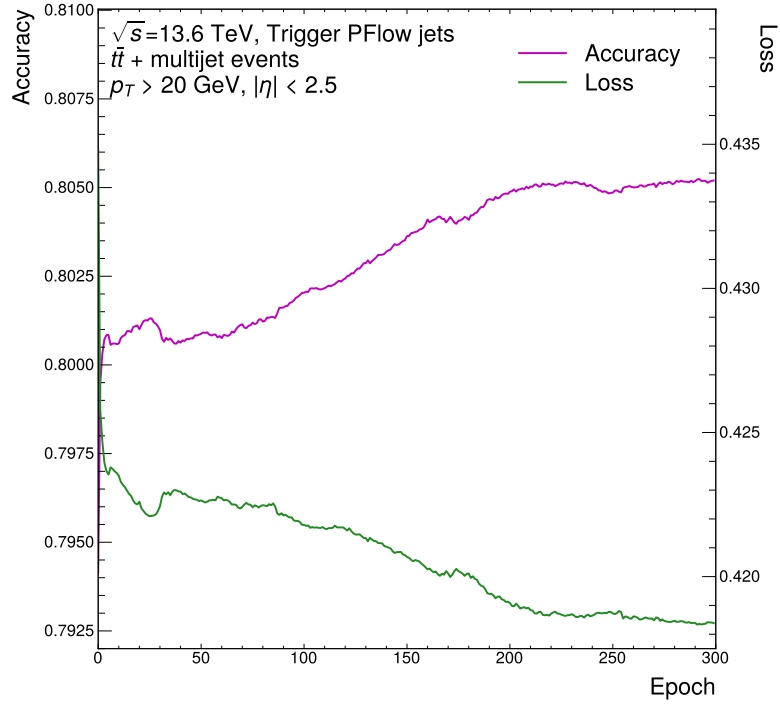


Figure 3.5: Training loss and accuracy evolution over the 300 training epochs of the DL1dbb algorithm.

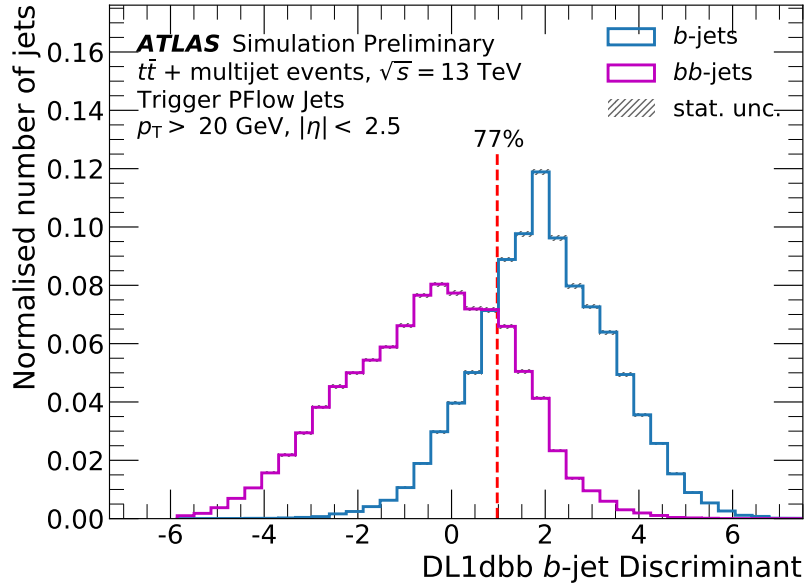


Figure 3.6: Distributions of the b -jet discriminator for the DL1dbb algorithm, of single- b -jets and bb -jets in simulated $t\bar{t}$ and QCD multijet events. All jets are HLT Particle Flow jets, and are required to have $p_T > 20$ GeV and $|\eta| < 2.5$. The red dashed vertical line represents the working point at 77% single- b -jet efficiency. The figure is published in [3].

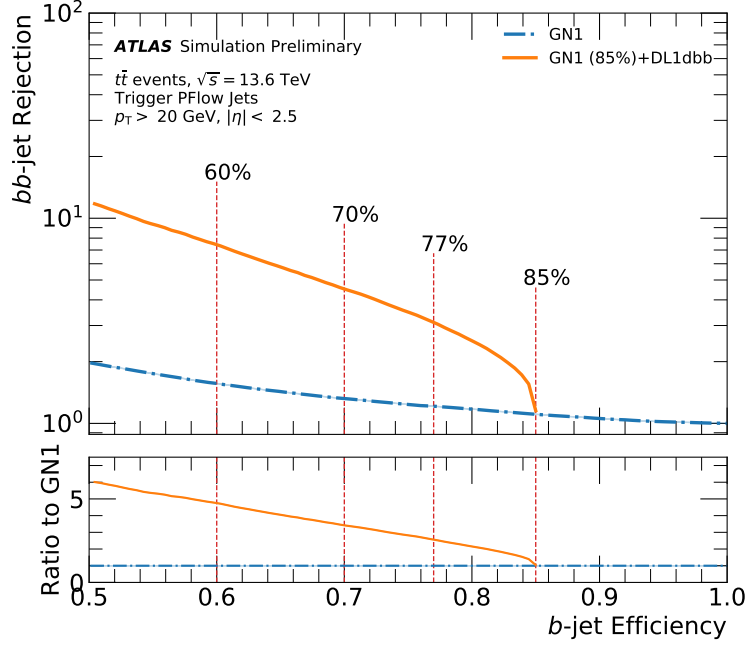


Figure 3.7: The rejection of bb -jets as a function of b -jet efficiency, of the baseline b -jet trigger b -tagging algorithm GN1, and the combined tagging scheme of a fixed GN1 working point at 85% b -jet efficiency in conjunction with varying working points of DL1dbb. The rejection is evaluated on trigger Particle Flow jets with $p_T > 20$ GeV and $|\eta| < 2.5$ in simulated $t\bar{t}$ events. The red dashed vertical lines represent the 4 points of overall b -jet efficiencies from combining the GN1 and DL1dbb working points. The figure is published in [3].

3.4.1 Working Point Scan

The optimal combinations of working points is searched for with a trigger chain used in the ATLAS Run-2 non-resonant $HH \rightarrow 4b$ analysis [13], shown in Figure 3.8. The analysis targets a final state consisting of 4 resolved b -jets produced from the decay of 2 Higgs bosons. A 3b1j trigger chain is tested, which requires at least 4 jets at various p_T thresholds, and at least 3 jets passing the b -tagging requirement. The rejection of multijet events containing fewer than 3 single- b -jets is plotted as a function of the efficiency of genuine non-resonant $HH \rightarrow 4b$ signal events containing at least 4 single- b -jets for this chain. The b -tagging threshold of the original trigger chain is set by GN1 only, where the 6 defined working points are at b -jet efficiency of 70%, 75%, 80%, 82% and 85%. The coloured stars represent the background rejection achieved by requiring 3 jets tagged by GN1 only at these working points. Then, for each of these GN1 working points, the background rejection is evaluated by requiring 3 jets tagged by a combined requirement of GN1 and various working points of DL1dbb, represented by the solid coloured lines. The combinations of working points that achieve

a higher background rejection than the GN1-only b -tagging requirement at a given $HH \rightarrow 4b$ signal efficiency are searched for. Three of such combinations are found, and the ratios of their background rejections to that achieved by the GN1-only b -tagging requirement are plotted in the bottom panel in Figure 3.8. The details of these working point combinations are summarised in Table 3.4.

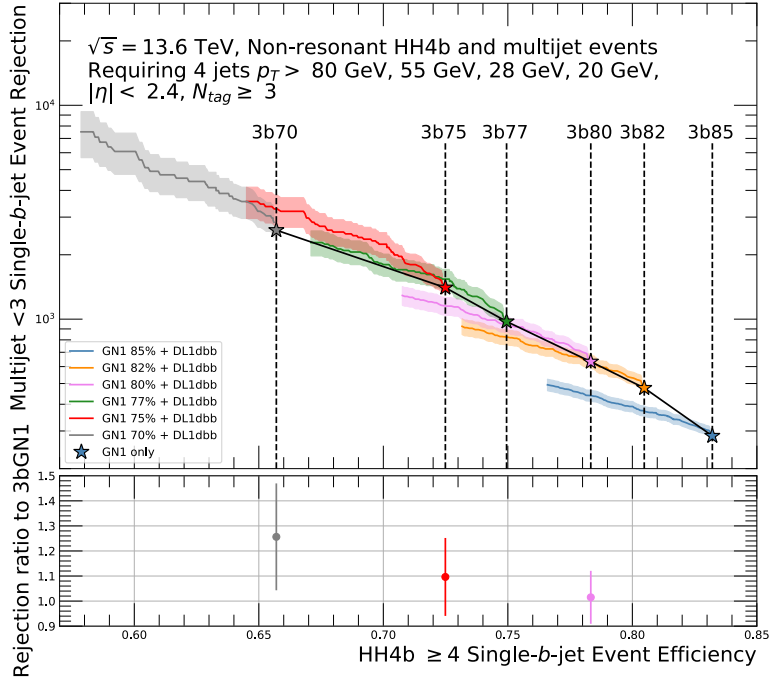


Figure 3.8: The rejection of background events consisting of multijet events with fewer than 3 single- b -jets as a function of the efficiency of signal events consisting of non-resonant $HH \rightarrow 4b$ events with at least 4 single- b -jets, evaluated for a 3b1j trigger chain requiring 4 jets and at least 3 to pass the b -tagging requirement. The coloured stars represent the background rejections where the b -tagging requirement consists only of GN1 at the currently defined working points. The solid coloured lines represent background rejection where fixed working points of GN1 and various working points of DL1dbb are combined. The coloured bands show their binomial errors. The vertical dashed lines point to the overall $HH \rightarrow 4b$ signal efficiencies given by the GN1-only b -tagging requirement, namely 3b-(working point). The bottom panel shows the ratios between background rejections of the combined to the GN1-only tagging scheme at 3 different $HH \rightarrow 4b$ signal efficiencies, where the background rejection of the combined tagging scheme is higher.

3.4.2 2-dimensional Discriminant Visualisation

When using two b -tagging algorithms in conjunction, it is useful to visualise their discriminants in a 2-dimensional plane, such as in Figure 3.9 with the b -jet discriminant of DL1dbb on the x -axis and of GN1 on the y -axis. The GN1 working points corresponding to 70% and 75% b -jet efficiencies are

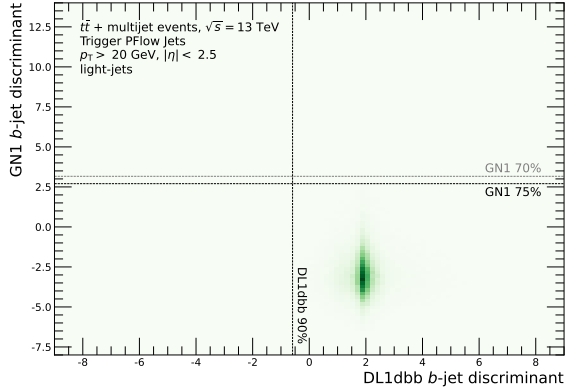
GN1-only	Combined tagging scheme	
GN1	GN1	DL1dbb
80%	82%	96%
75%	77%	96%
70%	75%	90%

Table 3.4: Working points of the GN1-only and the combined tagging scheme that yield the same non-resonant $HH \rightarrow 4b$ signal efficiencies

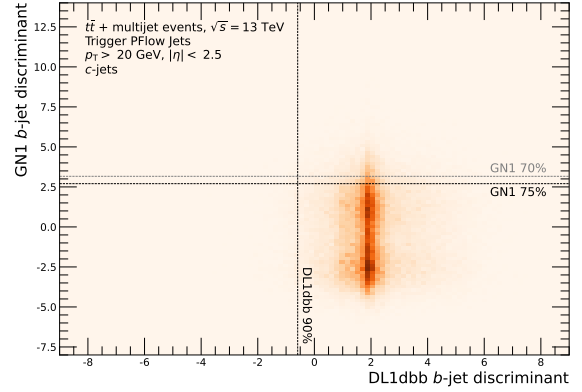
shown as the black horizontal dashed line, and jets above it are tagged as b -jets. The DL1dbb working point corresponding to 90% b -jet efficiency is shown as the vertical dashed line, and jets to the right of it are tagged as single- b -jets. The jets in the upper right quadrant pass both working points. The combination of the 75% working point of GN1 and 90% working point of DL1dbb is found to have a higher multijet background rejection for $HH \rightarrow 4b$ signal in the working point optimisation scan described in the previous section. The overall b -tagging efficiency from this combination is roughly equal to using the 70% working point of GN1 only, represented by the grey dashed line. This visualisation illustrates that GN1 is proficient at rejecting c -jets and light-flavour jets, as there are very few of them in the upper right quadrant. However it struggles to reject bb -jets. Whereas DL1dbb struggles to reject c -jets and light-flavour jets as it is not trained to do so, but it is better at rejecting bb -jets. With the combination of the two algorithms, a larger fraction of bb -jets is rejected, while trading for a slightly higher contamination from c -jets and none from light-flavour jets.

3.5 Implication in the b -jet Trigger

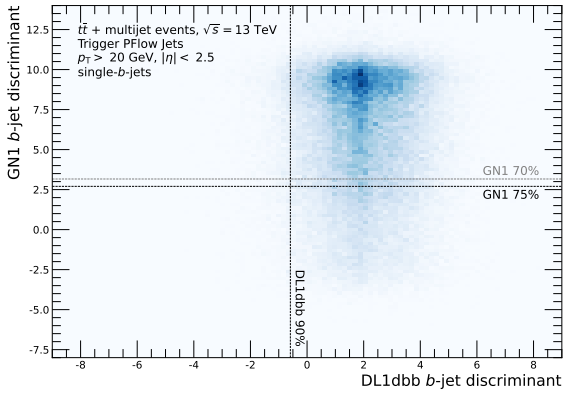
Aside from a higher rejection of background events at the trigger level for physics analyses, another important effect of DL1dbb is the potential reduction of trigger readout rates. At ATLAS, special datasets called enhanced bias data are taken specifically to study trigger rates of various algorithm development at the HLT, where selection bias from the triggering is removed using event weights [74]. In order to encompass a wide range of kinematic regions, events are collected using a mixture of L1 requirements to emphasise events with high p_T and object multiplicities. The trigger rates of the 3b1j trigger chain used in the non-resonant $HH \rightarrow 4b$ analysis studied in the previous section



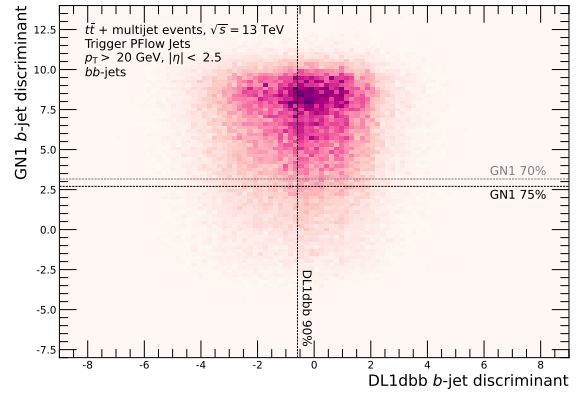
(a)



(b)



(c)



(d)

Figure 3.9: 2-dimensional visualisation of the GN1 and DL1dbb b -jet discriminants evaluated in light-jets (green), c -jets (orange), single- b -jets (blue) and bb -jets (purple). The black dashed lines represent the thresholds on the two discriminants corresponding to the GN1 and DL1dbb working points at 75% and 90% b -jet efficiencies respectively. The jets in the upper right quadrant pass their combined b -tagging requirement. The grey dashed lines represent the GN1 discriminator threshold at the 70% working point, which roughly equals the overall b -jet efficiency from the combined tagging scheme.

are estimated, with the three optimal working point combinations found to have higher multijet background rejection. The estimated rates are compared with the rates from the same chain but with GN1-only b -tagging requirement, at various GN1 working points, as a function of their b -jet efficiency, shown in Figure 3.10. A reduction of rate is estimated for the working point combination of (GN1, DL1dbb) at (77%, 96%) and at (75%, 90%), compared to the closest b -jet efficiency given by a GN1 working point. The combination of (GN1, DL1dbb) at (82%, 96%) yields a higher trigger rate. The reduction of rates occurs at GN1 working points with lower b -jet efficiency as a result of the trade-off between the rejection of bb -jets and other background jet flavours. This is evident in the b -jet discriminant distributions of GN1 in Figure 3.3, where, between 75% and 80% b -jet efficiencies, the fraction of c -jets increases more rapidly than bb -jets. Whereas at lower b -jet efficiencies, the increase in fraction of bb -jets is more significant compared to c -jets and light-flavour jets, where DL1dbb would have a larger impact on the background rejection as well as trigger rate reduction.

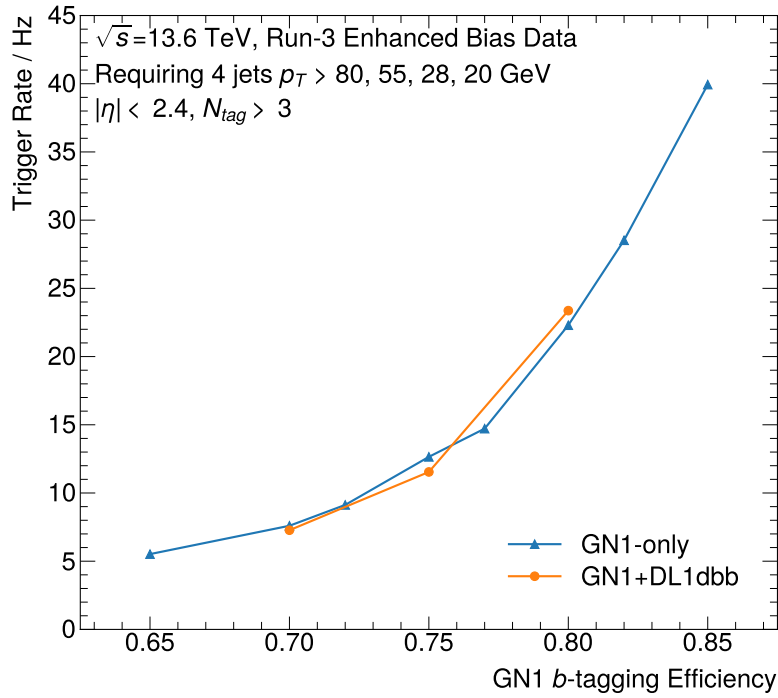


Figure 3.10: The trigger rates of the 3b1j trigger chain used in the Run-2 non-resonant $HH \rightarrow 4b$ analysis, evaluated on Run-3 enhanced bias data. The rates are compared between b -tagging requirement with GN1 algorithm only, and with GN1 and DL1dbb together at three optimal working points found to have higher multijet background rejection.

3.6 Conclusions and Outlooks

This study shows that the addition of bb -jet rejection in the current b -jet triggers can increase the rejection of multijet background arising from $g \rightarrow b\bar{b}$ splitting for analyses such as non-resonant $HH \rightarrow 4b$ and in turn reduce readout rates. This is achieved by training a dedicated algorithm to classify single- b -jets and bb -jets, and applying it in conjunction with the baseline b -tagging algorithm in the b -jet triggers. However, the gain from rejecting bb -jets is only found at low b -jet efficiencies, where the contamination from bb -jets is more significant compared to c -jets and light-flavour jets. This is evident in the estimation of multijet rejection and the rates on enhanced bias data for the 3b1j trigger chain, where improvements are only seen for working points corresponding to lower b -jet efficiencies. As a result, the DL1dbb algorithm is not deployed for the ATLAS Run-3 data-taking. This study however demonstrates that there is a non-negligible amount of contamination from bb -jets at the trigger level, which the current baseline b -tagging algorithms identify as genuine single- b -jets. The kinematic differences of these two jet flavours are significant enough to provide separation between them using b -tagging algorithms. With the new family of algorithms based on GNN and transformer architecture [75], it might be possible to add bb -jets as an additional jet flavour class at a minimal cost of the rejection of c -jets and light-flavour jets which is a limitation for the DL1d family of algorithms. These algorithms also achieve impressively high rejections of c -jets and light-flavour jets, which would increase the relative impact of bb -jets contamination. As b -tagging algorithms advance further to handle increasingly complex tasks, it is worth re-visiting the rejection of bb -jets in the future.

4 | The Search for Triple Higgs Boson Production in the 6-*b*-quark Final State

They would certainly spend a lot of time trying to prove it's a bug.

— Patrick Dougan

4.1 Introduction

A core focus of the physics programme at the ATLAS experiment is to discover new physics. This goal is generally approached in two ways, via the precise scrutiny of parameters predicted by the Standard Model (SM), and via direct searches for signatures of beyond the Standard Model (BSM) physics. The simultaneous production of three Higgs bosons, also known as *tri-Higgs* (HHH) production, provides an opportunity for both approaches.

Among the numerous important SM parameters is the Higgs self-coupling, which describes the strength at which Higgs bosons interact with each other. The SM HHH production is dependent on the Higgs self-coupling, such that measuring its rate, or *cross-section*, allows a direct probe of the *trilinear* (λ_3) and the *quartic* (λ_4) Higgs self-coupling constants. The dependence of SM HHH production on these coupling constants are illustrated in Figure 4.1. Their deviations from the SM expectations are characterised by *coupling modifiers*, which are expressed as their ratios to the SM predictions:

$$\kappa_3 = \frac{\lambda_3}{\lambda_3^{SM}}, \kappa_4 = \frac{\lambda_4}{\lambda_4^{SM}} \quad (4.1)$$

Note that the SM HHH production can also occur via the "pentagon" diagram in Figure 4.1(d) which does not depend on λ_3 or λ_4 , and has a larger cross-section.

Beyond the SM, a family of theories in the extended scalar sectors also predicts HHH production at an enhanced cross-section. The HHH whitepaper in [76] provides an overview of the potential

BSM scenarios and feasible channels that explore HHH production for current and future searches. The Benchmark BSM model considered in this analysis is the "Two Real Scalar Model" (TRSM), in particular the Benchmark Point 3 (BP3) in [77]. This particular Benchmark Point is chosen as its TRSM parameters maximise the HHH production cross-section through a "cascade decay", shown in Figure 4.1(a). Here, the scalar potential of the SM is extended by two additional real scalar fields, which manifests as two mass eigenstates that can be observed as heavy scalar particles X and S . If the masses of both particles are sufficiently high, that is if both are produced on-shell, three SM-like Higgs bosons can be produced via the "cascade decay" at an enhanced production cross-section of around 50 fb. The on-shell X results in a distinct peak, also referred to as a resonance, in the reconstructed tri-Higgs invariant mass distribution. If only the heavier scalar particle X is on-shell while the lighter scalar particle S is off-shell, this can still lead to non-resonant HHH production at a sizeable cross-section. However if both scalar particles are off-shell, the production cross-section becomes heavily suppressed. For this reason, only the cases where both X and S are on-shell or when only X is on-shell are targeted by this analysis.

The results of the resonant and non-resonant interpretation can also be interpreted in an alternative BSM model, the "simple model for dark matter and CP violation" (DM-CPV) [78], using benchmarks 1 and 3 which maximise HHH production. This model predicts dark matter as a vector-like dark fermion that interacts with the SM through an electric singlet. Negligible kinematic differences are found between HHH events produced from the TRSM and DM-CPV models, making the results interpretable in both models by only generating events using the TRSM.

HHH production can also occur via the production of generic heavy spin-0 bosons X and S that have masses larger than what is allowed by perturbative unitarity in BP3 of the TRSM. These scenarios are included with both particles having either a wide decay width, or a narrow decay width that is above the detector resolution. Only resonant production via the "cascade decay" is considered to simplify the simulation process and to avoid large contributions from non-resonant diagrams at leading order in the TRSM.

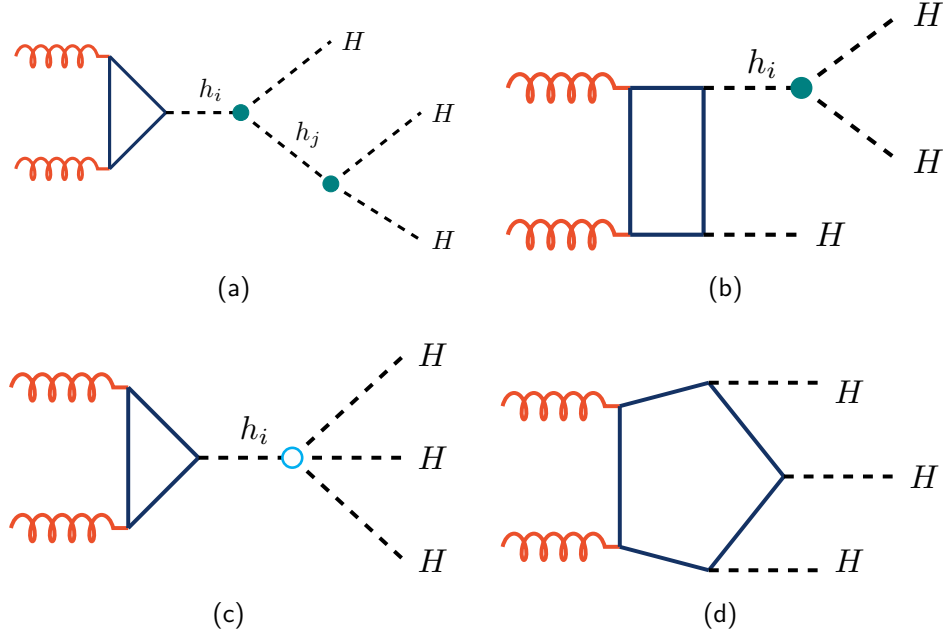


Figure 4.1: Leading-order Feynman diagrams of tri-Higgs production via gluon-gluon fusion. For SM-like production, the h_i and h_j represent SM Higgs boson H , and the solid circle in (a) and (b) represents the trilinear Higgs self-coupling λ_3 , and the open circle in (c) represents the quartic Higgs self-coupling λ_4 . For BSM tri-Higgs production, the h_i and h_j represent heavy scalar particles X or S . These diagrams are taken from [1].

4.1.1 Physics Interpretations

The analysis interprets the observed data in three categories, the non-resonant, the resonant, and the heavy resonant interpretations.

- **Non-resonant:** targets SM-like HHH production where the trilinear and quartic Higgs self-coupling constants are modified with respect to the SM expectations. It also includes HHH production from the TRSM where the lighter scalar particle S is off-shell, i.e. $m_S \leq 250$ GeV.
- **Resonant:** targets tri-Higgs production via the cascade decay of resonances in the TRSM, where both scalar particles are sufficiently heavy.
- **Heavy resonant:** targets HHH production via generic heavy scalar particles with masses higher than the allowed values in the TRSM. To increase re-interpretability, each mass hypothesis in this interpretation considers both a narrow and a wide decay width scenario, where they correspond to 1% and 20% of the respective scalar particle masses.

The mass ranges of the scalar particles predicted by the BSM models in each of the physics inter-

pretation are summarised in Table 4.1.

	Non-resonant	Resonant	Heavy resonant
m_X [GeV]	$325 \leq m_S \leq 575$	$400 \leq m_S \leq 550$	$600 \leq m_S \leq 1500$
m_S [GeV]	$200 \leq m_S \leq 250$	$250 < m_S \leq 350$	$275 \leq m_S \leq 1000$

Table 4.1: Table summarising the mass ranges of the scalar particles X and S predicted by the BSM theories in each of the physics interpretation considered in the analysis.

The categorisation is chosen according to the production topology of the targeted signals, which leads to significantly different event kinematics. For example, the distribution of the reconstructed invariant mass of the three Higgs bosons (the tri-Higgs invariant mass m_{HHH}) of the events containing at least 6 b -tagged jets are shown in Figure 4.2(a) at several representative scalar particle masses. For the resonant interpretation, the TRSM signal events tend to have a narrow peak in the m_{HHH} distribution which corresponds to the mass of the heavier scalar particle m_X . At low m_X , and therefore low m_{HHH} , the Higgs bosons produced are generally less boosted. As a result, the b -jet pairs from their decay are also less collimated and have larger angular separations. This is evident in the distribution of the angular separation of the b -jet pair from the leading Higgs boson candidate decay (ΔR_{H1}) shown in Figure 4.2(b), where the resonant TRSM signal events have the largest b -jet pair separation. For the non-resonant interpretation, the TRSM signal events tend to have higher values in m_{HHH} due to the absence of an on-shell scalar particle S . Their m_{HHH} distribution is also similar to the SM HHH signal. Correspondingly, their ΔR_{H1} also have smaller values as the Higgs boson candidates are more boosted, producing more collimated b -jet pairs as they decay. For the heavy resonant interpretation, the heavy-resonance signal events tend to have the narrowest peak due to the heavy scalar particle m_X . The signals generated with a narrow decay width also have a narrower peak than ones generated with a wide decay width. The ΔR_{H1} from these signals have the smallest values, as the Higgs boson candidates produced from heavier scalar particles are strongly boosted, decaying into collimated b -jet pairs.

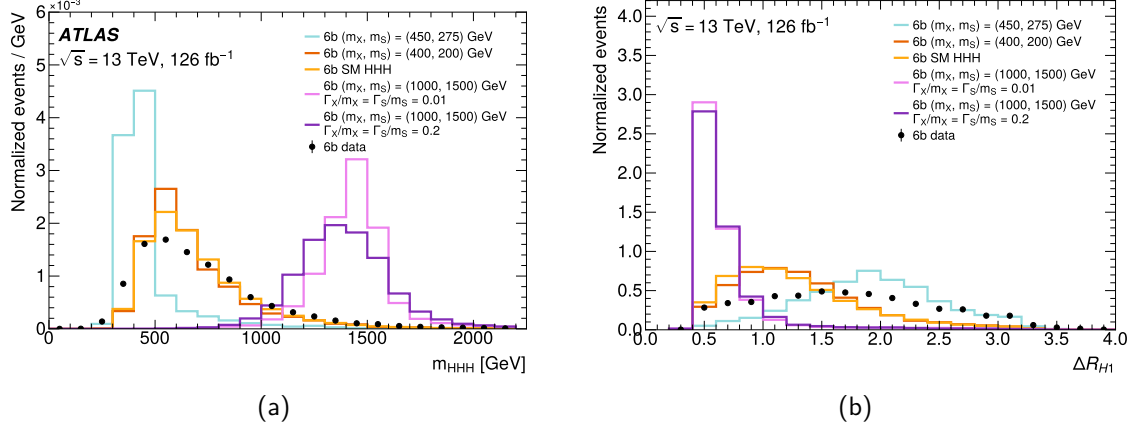


Figure 4.2: The reconstructed tri-Higgs invariant mass m_{HHH} (a) and the angular separation between jets used to construct the leading Higgs boson candidate ΔR_{H1} 4.2(b) of representative signal samples in each physics interpretation. For the resonant interpretation, a TRSM signal where both X and S are on-shell is shown in light blue. For the non-resonant interpretation, a TRSM signal where S is off-shell is shown in dark orange, and the SM HHH signal is shown in light orange. For the heavy resonant interpretation, two generic resonance signals are shown with a narrow and a wide decay width for the same m_X and m_S values in pink and purple respectively. The observed $6b$ data is shown as black data points. All histograms are normalised to have unit area for shape-only comparison. All events pass the selection criteria detailed in Section 4.3. Figure (a) is published in [1].

4.1.2 Analysis Strategy Overview

The search in each physics interpretation follows the same analysis strategy. Recorded data events are selected by a set of triggers detailed in Section 4.3.1, and required to pass the selection criteria detailed in Section 4.3 to ensure a high signal purity. The selected events are divided into three categories, based on the number of b -tagged jets (b -jet multiplicity), including a $6b$ Signal Region which contains the most signal-like events. Events in the Signal Region are classified by Deep Neural Networks (DNNs) into signal and background classes. A pair of DNNs is trained on the relevant signals and SM background events for each physics interpretation separately, documented in Section 4.4. The dominant SM background in all interpretations is multijet production arising from Quantum Chromodynamics (QCD), and is estimated using only observed data events (data-driven) instead of simulated events. The method is described in Section 4.5, and is based on the extrapolation across different b -jet multiplicities. The DNN for each interpretation is carefully designed such that this extrapolation is smooth. Finally, the DNN score is used as the main discriminant for statistical analysis in the form of a profile likelihood fit described in Section 4.8. Systematic

uncertainties arising from various sources such as the detector, theory, and background estimation are accounted for using nuisance parameters in these fits, detailed in Section 4.6. In particular, the difference between simulated and data events arising from the trigger conditions is carefully calibrated, using a procedure described in Section 4.7. Using a profile likelihood fit, upper limits on the various production cross-sections considered in each physics interpretation are placed and shown in Section 4.9. For the SM-like HHH production in the non-resonant interpretation, a dedicated statistical setup is used to constrain κ_3 and κ_4 values simultaneously, detailed in Section 4.8.4 and 4.9.2.

4.2 Data and Monte Carlo Samples

Data corresponding to 126 fb^{-1} of proton-proton (pp) collisions at a centre-of-mass energy of $\sqrt{s} = 13 \text{ TeV}$ collected between 2016 and 2018 during the Run-2 data-taking by the ATLAS detector is used in this analysis. The b -jet triggers used to select the recorded data events in each year are documented in Section 4.3.1.

Signal samples for processes considered in all analysis interpretations are simulated using MADGRAPH 2.9.5 [79] at leading-order in the strong coupling constant using the NNPDF3.0NLO parton distribution function (PDF) set [80], with the A14 set of tuned parameters [81]. The simulated Higgs boson is taken to be SM-like, with a mass of 125 GeV and a total width of 4.088 MeV [82]. The branching ratio of $H \rightarrow b\bar{b}$ is taken to be 0.582. Parton showering is handled with PYTHIA 8.245 [70]. At generation level, a requirement of at least 4 b -tagged jets with $p_T > 25 \text{ GeV}$ is placed on the events.

For the SM HHH signal, the events are generated with $(\kappa_3, \kappa_4) = (1, 1)$, and the contribution and interference between all diagrams in Figure 4.1 are taken into consideration. The Higgs bosons are decayed exclusively to $b\bar{b}$. For SM-like HHH production with different κ_3 and κ_4 values, their impacts on the event kinematics are accounted for using a reweighting method [83]. This method allows the $(\kappa_3, \kappa_4) = (1, 1)$ to be reused to generate other signal samples with varying κ_3

and κ_4 values by reweighting it with a weight proportional to the ratio of their matrix elements. The theory predicted SM HHH production cross-section via ggF at a centre-of-mass energy of $\sqrt{s} = 13$ TeV is taken as $\sigma_{HHH}^{SM} = 0.079_{-0.013}^{+0.012}$ fb at next-to-next-to-leading-order (NNLO). This value is extrapolated from a value calculated at a higher centre-of-mass energy in [84].

The resonant and non-resonant interpretations target BP3 of the TRSM. The corresponding signal samples are generated for the 2 neutral CP-even scalars X and S , with masses such that $m_X \geq m_S \geq m_H$. Contributions from the "cascade decay" diagram in Figure 4.1(a) as well as all possible $gg \rightarrow HHH$ diagrams at leading-order and their interference are considered. The branching fractions of $X \rightarrow SH$ at 45% and $S \rightarrow HH$ at 70% are taken from [77]. The Higgs bosons are decayed inclusively using PYTHIA 8.245 [70]. Signal samples of 27 (m_X, m_S) points that are within the perturbative unitarity bounds of the TRSM are generated. Their mass ranges are $325 \leq m_X \leq 575$ GeV and $200 \leq m_S \leq 350$ GeV, which include cases where both X and S are on-shell leading to resonant production ($m_X > m_S + m_H$ and $m_S > 2m_H$), and where S is off-shell leading to non-resonant production ($m_S \leq 2m_H$).

To generate signal samples of generic heavy scalars targeted by the heavy-resonant interpretation, BP3 of the TRSM is used as a baseline model, but only the "cascade decay" diagram is considered. The narrow and wide decay widths of the resonance particles are defined as being 1% and 20% of their masses respectively.

The detector response and interactions are simulated using fast parametrised simulation of the calorimeters [85], and GEANT4 for other sub-detectors. The physics objects detailed in Section 4.3.2 in data and simulated events are reconstructed using the same software. Pileup is modelled by pp inelastic scattering events generated with PYTHIA 8.186 [86] that uses the NNPDF2.3LO PDF sets [87] with A3 tuning [88].

4.3 Event Reconstruction and Selection

The chosen set of triggers used to select the recorded data events are documented in Section 4.3.1. The definitions and reconstruction of the physics objects in the events are described in Section 4.3.2, before the events are separated into the categories outlined in Section 4.3.3. A selection of jets in the events in each category are further paired into Higgs boson candidates using a method described in Section 4.3.4 to reconstruct kinematic variables specific to these candidates.

4.3.1 Trigger

Physics objects are first reconstructed at the trigger level using methods detailed in Section 2.4, before selected by the relevant triggers. For this analysis, events are selected using a b -jet trigger chain for each data-taking year, summarised in Table 4.2. These b -jet trigger chains are chosen to maximise the efficiency of selecting $6b$ TRSM signal events. They are evaluated to be above 90% for $6b$ signal events across all the simulated scalar particle masses, since the probabilities of these events containing 2 or 3 b -tagged jets required at the trigger level would be high given that at least 6 b -tagged jets are required at later reconstruction.

The jet p_T requirements specified in these triggers are in the low p_T region where the per-jet trigger efficiency is usually below 100%. This region is often referred to as the *turn-on* region, where larger discrepancies in per-jet trigger efficiencies are typically seen between simulated and data events. To correct for these differences, the trigger efficiencies are calibrated using a dedicated method detailed in Section 4.7.

4.3.2 Object Definitions

Events passing the trigger requirements are further reconstructed to higher accuracies using offline reconstruction. Particle flow (PFlow) jets with $R = 0.4$ are reconstructed using methods described in Section 2.5. To reject jets originating from pileup events, jets are required to have $p_T > 20$ GeV and $|\eta| < 2.5$, and to have passed a jet vertex tagger (NNJVT [52]). PFlow jets are then evaluated by a DNN-based b -tagging algorithm (DL1d) [89]. The working point corresponding to

Year	Type	Trigger chain name (Conditions at HLT)
2016	2b2j	HLT_2j35_bmv2c2060_split_2j35_L14J15.0ETA25 (2 jets b -tagged with 60% WP and $p_T > 35$ GeV, and 2 additional jets with $p_T > 35$ GeV)
2017	3b1j	HLT_3j15_gsc35_bmv2c1070_split_j15_gsc35_boffperf_split_L14J15.0ETA25 (3 jets b -tagged with 70% WP and $p_T > 35$ GeV, and 1 additional jet with $p_T > 35$ GeV)
2018	2b2j	HLT_2j35_bmv2c1060_split_2j35_L14J15.0ETA25 (2 jets b -tagged with 60% WP and $p_T > 35$ GeV, and 2 additional jets with $p_T > 35$ GeV)

Table 4.2: Table summarising the b -jet trigger chains used to select events for the analysis in data-taking years 2016, 2017 and 2018 and the conditions at the HLT level in each chain. To interpret the trigger chain names, the conditions at the HLT level follow the prefix "HLT_" and "L1_" for the L1 level. For example, 2j35 refers to the condition requiring 2 jets with $p_T > 35$ GeV before calibration, while some chains specify the jet p_T requirements after global sequential calibration [56] such as "gsc35" which corrects the p_T requirement to be above 35 GeV. This is followed by the b -tagging requirement, where "bmv2c10" and "bmv2c20" refer to the names of Boosted Decision Tree b -tagging algorithms used in the ATLAS Run-2 data-taking periods, followed by the working point used. "split" specifies the type of jets used in the trigger. The prefix "L1_" is followed by the number of jets required at a given p_T threshold and the requirements in absolute pseudorapidity $|\eta|$. All three triggers used require 4 jets with $p_T > 15$ GeV and $|\eta| < 2.5$ at L1. An event passes the trigger if it satisfies conditions at both L1 and HLT.

77% b -jet efficiency is used, and the corresponding mis-tagging rate for c -jets is 15.5%, and 0.38% for light-flavour jets. For a final state that corresponds to 6 b -jets, the signal efficiency at this b -jet efficiency is a mere 20.9%. However, a higher b -jet efficiency comes with higher mis-tagging rates for c -jets and light-flavour jets, common in the dominant background process which is QCD multijet production. Indeed, a study using the working point of DL1d corresponding to the 85% b -jet efficiency shows an overall reduction in the sensitivity to $6b$ resonant TRSM signals.

Since the analysis uses both simulated and recorded data, the difference in b -tagging efficiencies between them are accounted for. The b -tagging efficiencies and light-jet mis-tagging efficiencies in simulated data are corrected to match the recorded data [59]–[61]. To account for energy lost in the semi-leptonic decay of the b -hadrons, a correction of the momentum of the b -hadron is applied [90]. The muon candidates used to derived this correction are reconstructed using methods described in Section 2.5.

4.3.3 Event Selection and Categorisation

Additional criteria are placed in order to select the most signal-like events. Events are required to have at least 6 PFlow jets, where at least 4 with $p_T > 40$ GeV and $|\eta| < 2.5$. Orthogonally, at least 4 jets are required to be b -tagged by DL1d.

For events passing both the trigger and the above event selection, they are categorised into three orthogonal regions according to their b -jet multiplicity, summarised in Table 4.3. The $6b$ region is defined as the Signal Region (SR) as it is enriched in the targeted signal events. The $5b$ and $4b$ regions contain events mostly arising from the dominant background process, QCD multijet production. There is a non-negligible amount of $6b$ signal events in the $5b$ region, where one of the b -jets is mis-identified by the b -tagging algorithm. The amount of contamination in various b -tag regions is quantified by calculating the Z significance detailed in Appendix D. To account for this signal contamination in the statistical analysis, the parameter of interest of the fit - the signal strength, is correlated across the $6b$ and the $5b$ region, detailed in Section 4.8. The amount of $6b$ signal events in the $4b$ region however is insignificant, and is therefore not dealt with the same treatment.

b -tag region	$6b$	$5b$	$4b$
b -jet requirement	≥ 6 b -jets	$= 5$ b -jets	$= 4$ b -jets
Function	Signal region	Background estimation & validation	

Table 4.3: Table summarising the three regions defined in the analysis, their b -jet requirements, and their functions in the analysis.

An event display of a candidate $HHH \rightarrow 6b$ event recorded in 2018 passing the trigger and event selections is shown in Figure B.1.

4.3.4 Jet Pairing

Once an event is categorised according to its b -tag multiplicity, 6 jets are selected to be assigned to 3 Higgs boson candidates (*jet pairing*). This allows the individual Higgs candidates as well as their kinematics to be reconstructed, which are used as powerful input features to the DNNs for signal background discrimination, described in Section 4.4. For events with exactly 6 b -jets, all 6 constitute the ones used for jet pairing. For events with < 6 b -jets, the 6 jets used for jet pairing consist of all available b -tagged jets and non- b -tagged jets sorted in descending p_T until 6 jets have been selected. For events with > 6 b -jets, the 6 leading b -tagged jets are chosen.

The task of assigning 6 jets to 3 Higgs boson candidates is a complicated one, as 15 unique combinations arise from $\binom{6}{2} = 15$. In previous multi-Higgs analyses targeting multiple b -jets in the final state, namely $HH \rightarrow 4b$, where 4 jets are assigned to 2 Higgs candidates, there are 6 unique combinations, and the jets with the lowest angular separation (ΔR) are paired together. This however is not suitable in a 6-jet scenario, as the jets have a higher chance of overlapping. Two different approaches are explored for this task, one using a DNN to classify pairs of objects as correct or incorrect pairings, and one using a mass-based method. The latter is used in the final analysis as it yields higher pairing efficiency over all the signal samples. It assigns the 6 jets to minimise the quantity:

$$|m_{H1} - 120 \text{ GeV}| + |m_{H2} - 115 \text{ GeV}| + |m_{H3} - 110 \text{ GeV}| \quad (4.2)$$

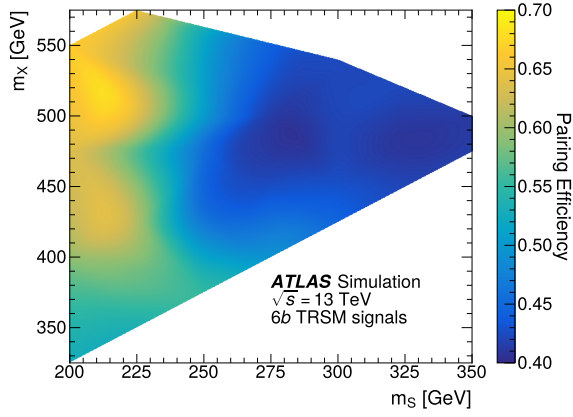
where m_{H1} , m_{H2} and m_{H3} are the invariant mass of the leading, sub-leading and least-leading Higgs boson candidates formed from the pairing. This method effectively forms pairs of jets with an invariant mass close to the mass of the SM Higgs boson at 125 GeV. The cascading mass values are to account for the detector resolution that is dependent on the jet p_T , and are taken as the mean masses of the Higgs bosons in the simulated $6b$ TRSM signal events. The pairing efficiency in the simulated signal samples is evaluated as percentage of events with the correct pairing over events where all 6 b -jets are each matched to a b -hadron at truth level. They are evaluated for $6b$ signal events simulated at various scalar particle masses and at different (κ_3, κ_4) values, shown in Figure 4.3. The efficiencies are higher for heavy-resonant signal events, as the scalar particles have higher masses, resulting in more boosted Higgs bosons which decay into more collimated jets. This reduces the angular separation between the correct pairs of jets, and therefore reduces the ambiguity in the assignment. The pairing efficiencies for signal samples with various (κ_3, κ_4) values are higher for values that correspond to higher m_{HHH} containing more boosted Higgs boson candidates. This is a result of the interference between Feynman diagrams of SM-like HHH production, discussed in Ref. [15], where the four-momenta of the Higgs bosons and the m_{HHH} are sensitive to the interference between Feynman diagrams dependent on κ_3 .

4.4 Signal and Background Discrimination

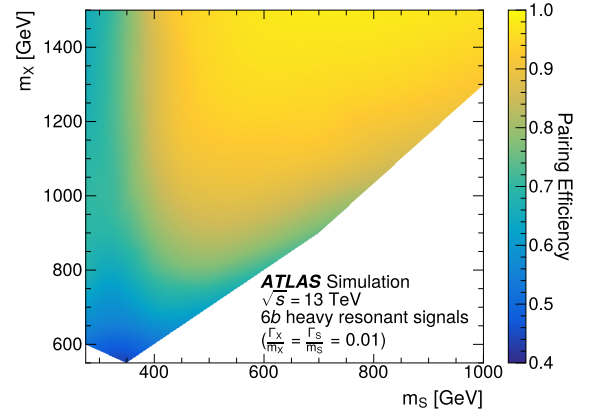
To separate signal events from SM background events, a pair of DNNs was trained to perform binary classification for each analysis interpretation separately. They are referred to as the ResDNN for the resonant interpretation, NonresDNN for the non-resonant interpretation, and the HeavyresDNN for the heavy-resonant interpretation. The output scores of the DNNs are used as the main discriminant in the statistical analysis.

4.4.1 Training Setup

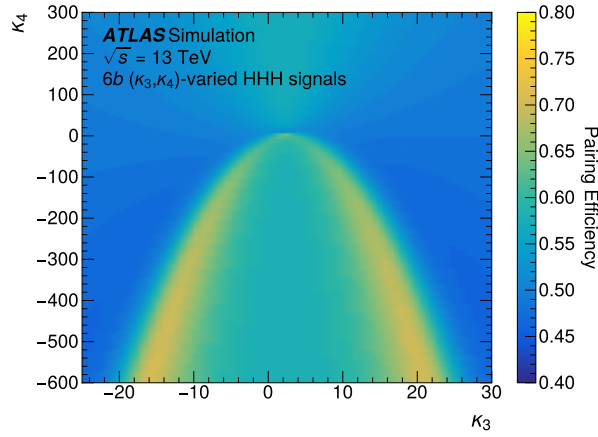
The DNNs are full-connected, feedforward deep neural networks implemented using Keras software, and trained with Tensorflow backend to perform binary classification. The two class labels are signal, representing the simulated $6b$ SM or BSM signal events in each interpretation, and background,



(a)



(b)



(c)

Figure 4.3: Pairing efficiencies for the simulated 6b TRSM signal samples (a), heavy-resonant signal samples with narrow decay width (b), and signal samples at various (κ_3, κ_4) values (c). The efficiencies between the simulated scalar mass values are interpolated using a cubic polynomial function. These figures are taken from [1].

representing the SM background. For the background class, $5b$ data events are used as the majority of them arise from QCD multi-jet production. For the signal class, in the non-resonant interpretation, only the SM HHH signal sample is used for the DNN training as it has sufficiently high statistics, and is representative of the kinematics of the non-resonant TRSM signal samples. For the signal class, in the resonant and heavy-resonant interpretations, events simulated for different scalar particle masses are combined to form a single training dataset, since the m_X and m_S are similar between samples. The signal samples with narrow and wide decay widths in the heavy-resonant interpretation are also combined as their kinematics are similar. A study is done where DNNs are trained for the narrow and the wide decay width signal samples separately and cross-applied with no significant drop in the performance observed, which confirms that the kinematics of different decay widths are similar enough for them to be combined in the training. Table 4.4 summarises the samples used to form the training dataset for the pair of DNNs in each interpretation. There are two sources of imbalanced statistics in the training: the different number of training events between the signal and background classes, and the different number of events between each signal sample in the resonant and heavy-resonant interpretations. To account for the first, a class weight is applied to each signal event:

$$w_{class} = \frac{N_{background}}{N_{signal}} \quad (4.3)$$

where $N_{background}$ is the total number of $5b$ data events and N_{signal} is the total number of signal events in each fold. To account for the latter in the resDNN and heavyresDNN training, a sample weight is calculated:

$$w_{sample} = \text{Norm} \left[\frac{w_i}{\sum_i^M w_i} * \sum_j^N \left(\sum_i^M w_i \right) \right] \quad (4.4)$$

where w_i is the event weight that represents the effective yield of each simulated signal event, taking into account the filter efficiency, luminosity, and theoretical production cross-section of the process. M is the total number of events in each mass hypotheses, and N is the total number of mass hypotheses in the resonant or heavy-resonant interpretation. The sample weights are normalised to have mean of 1:

$$\tilde{w}_{sample} = w_{sample} + (1 - \mathbb{E}[w_{sample}]) \quad (4.5)$$

The w_{class} and $\tilde{w}_{sample}$ are multiplied together and applied to each signal event to account for the two sources of statistics imbalance in the resDNN and heavyresDNN training. The nonresDNN training only uses the SM HHH signal sample, therefore only w_{class} is applied.

Model	Signal sample	Background sample
ResDNN	Combined $6b$ resonant TRSM signals	$5b$ data
NonresDNN	$6b$ SM HHH signal	$5b$ data
HeavyresDNN	Combined $6b$ narrow and wide decay width heavy-resonant signals	$5b$ data

Table 4.4: Table summarising the events that constitute the signal and background class in the training dataset for the pairs of DNNs trained for each analysis interpretation.

The DNNs perform binary classification between the signal and background class. For back propagation, the binary cross-entropy loss function in the following form is used

$$L = -(y \log(p) + (1 - y) \log(1 - p)) \quad (4.6)$$

where y is the truth label for each event (1 for the signal class and 0 for the background class), and p is the DNN output score for each event. Dropout regularisation, where random nodes in the hidden layers are omitted during the back propagation in each iteration, is used to prevent the model from overfitting to noise in the training dataset. The hyperparameters of each pair of DNNs are optimised through grid scans of a wide set of parameter combinations. Since the statistical analysis uses the DNN output distribution as the discriminating variable, the hyperparameters that give the best overall signal and background separation are selected. This separation is characterised by the Receiver Operating Characteristic (ROC) curve, used commonly in binary classification tasks, which represents how the truth positive rate (TPR) changes with respect to the false positive rate (FPR). The area under the ROC curve (AUC) quantifies the overall performance of the model over all values of the TPR, that is a perfect signal and background separation would yield an AUC of 1. The hyperparameter combination that yielded the highest AUC for the signal of each interpretation was selected, and shown in Table 4.5.

To prevent the models from learning features and noise specific to the training data (overtraining)

Parameter	ResDNN	NonresDNN	HeavyresDNN
Learning rate	0.01	0.005	0.005
Batchsize	128	256	256
Number of hidden layers	3	3	4
Number of nodes in each layer	24	24	24
Dropout rate	0.1	0.1	0.1

Table 4.5: Table summarising the optimised hyperparameters of the pairs of DNNs trained for each analysis interpretation.

while preserving training statistics, k-fold validation method is used where the entire dataset is split into k number of folds, and k number of models are trained. For each model, a fold is used as the validation dataset, and the remaining folds as the training dataset. For each interpretation, the combined signal and background dataset is split into 2 folds by splitting the events into even and odd event numbers, a number assigned sequentially to the events as they are recorded or simulated. The fold consisting of events with even event numbers becomes the training dataset for the even DNN model, and the fold consisting of events with odd event numbers becomes the validation dataset. The folds are then reversed for the training and validation of the odd DNN model. Thus, a pair of DNNs consisting of an even and an odd model is trained for each interpretation.

4.4.2 Input Feature Selection

Jet pairing enabled the reconstruction of individual Higgs boson candidates as well as their kinematic properties. Some of them are found to be particularly effective at separating signal and background events compared to low-level kinematics, such as ones at the jet level. In addition, event-level kinematics independent of jet pairing are also found to be effective. The input feature selection procedure is the same for each pair of DNNs across all analysis interpretations. To start with, a large number of kinematics are calculated, including ones associated with individual jets and with the overall events that are either dependent or independent of jet pairing. In particular, event shape variables such as sphericity and aplanarity are included to capture the correlation between jets, as they are sensitive to soft gluon emissions, common in a multijet final state. Sphericity measures the isotropy of the jet momenta. Aplanarity measures the amount of jet momenta lying outside of a

plane. They are derived from the momentum tensor an event M_{xyz} :

$$M_{xyz} = \sum_i \begin{pmatrix} p_{xi}^2 & p_{xi}p_{yi} & p_{xi}p_{zi} \\ p_{xi}p_{yi} & p_{yi}^2 & p_{yi}p_{zi} \\ p_{xi}p_{zi} & p_{yi}p_{zi} & p_{zi}^2 \end{pmatrix} \quad (4.7)$$

where i iterates over all jets in an event. Its eigenvalues are ordered and normalised such that $\lambda_1 > \lambda_2 > \lambda_3$. From them, the sphericity and the transverse sphericity are defined as

$$S = \frac{3}{2}(\lambda_2 + \lambda_3), S_{\perp} = \frac{2\lambda_2}{\lambda_1 + \lambda_2} \quad (4.8)$$

and aplanarity is defined as

$$A = \frac{3}{2}\lambda_3 \quad (4.9)$$

They are found to have good separation power shown in Figure 4.4, where 6 jets selected for reconstructing the Higgs boson candidates in HHH signal events tend to have a higher sphericity as they are more homogeneously distributed, and a higher aplanarity as they are less likely to be back-to-back along the longitudinal beam direction.

The input features for the DNN training for each analysis interpretation were selected by evaluating three metrics:

- **Feature importance** characterised by the Shapely value [91], which quantifies how impactful an input feature is on the model output. The higher the Shapely value, the more impactful it is.
- **Correlation** which indicates how much overlap there is in the information provided by a pair of input features.
- **Dependence on b -jet multiplicity** which needs to be minimised to ensure the smooth extrapolation in the DNN score across different b -tag regions when estimating the expected background in the $6b$ Signal Region.

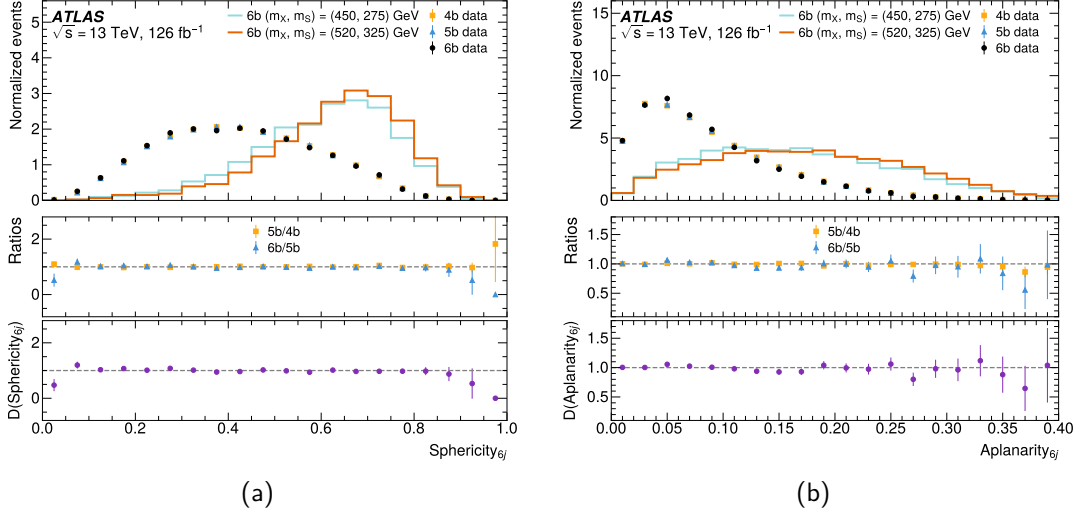


Figure 4.4: Distributions of sphericity (a) and aplanarity (b) evaluated on two representative $6b$ resonant TRSM signal samples and data in the $4b$, $5b$ and $6b$ regions. The bottom panel shows the double ratio D across the b -tag regions which is largely consistent with unity. These figures are published in [1].

First, input features are ranked by descending Shapley value to evaluate their impacts on the DNN outputs. However, Shapley values do not take into account the correlations between input features. For example two highly correlated or anti-correlated input features can provide repetitive information, but their Shapley values would be similar as they are equally discriminating. For two features with an absolute correlation > 0.7 , the one with a lower Shapley value is removed. This threshold value is chosen to strike a balance between retaining sufficient number of input features and the mitigation of overlapping information in them. In the case where a pair of input features are highly correlated in the signal sample, but highly anti-correlated in the background sample or vice versa, both input features are kept as the DNN would be able exploit the difference in the correlations between the two classes. The estimation of DNN score distribution in the expected $6b$ background relies on an extrapolation across different b -tag regions, therefore it is undesirable if the DNN score distribution varies drastically between these b -tag regions. To ensure this extrapolation is smooth, input features that are strongly correlated with the event b -jet multiplicity, such as the H_T , N_{jet} and m_{HHH} are excluded from the training. This is done by inspecting the input feature distributions in $5b$ and $4b$ data events, and selecting ones where the bin-by-bin ratios between $5b$ and $4b$ distributions do not vary significantly. Note that it is not required that these ratios are consistent with unity, as some kinematic differences between $5b$ and $4b$ data are expected due to the difference in the physics

processes they contain. A double ratio is defined as

$$D(I) = \frac{R(I)_{6b/5b}}{R(I)_{5b/4b}} \quad (4.10)$$

where I is an input feature, and $R(I)_{ib/jb}$ is the ratio of the feature between b -tag regions i and j . Its consistency with unity is an indication that the extrapolation of a given input feature across the various b -tag regions is smooth, and therefore the same extrapolation of the resulting DNN output score is valid. More detailed discussions on the validity of the background estimation can be found in Section 4.5.

The selected input features for resDNN, nonresDNN and HeavyresDNN and their definitions are summarised in Tables 4.6 and 4.7. Figure 4.5 shows the Shapley values of the selected input features for each model and their ranking.

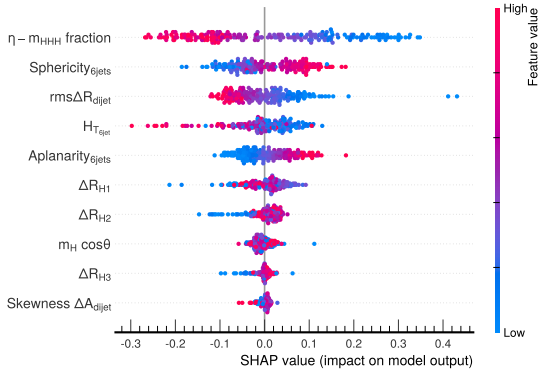
To offer some physical intuition of the most impactful input features, their distributions are shown in Figure 4.6. For the resonant interpretation, $\eta - m_{HHH}$ fraction measures the trade-off between the closeness to the longitudinal beam axis and p_T of each possible di-jet system, and the m_{HHH} of each event. The $5b$ data is enriched in QCD multijet background events which produce collimated jets from small-angle scattering that are close to the longitudinal beam axis, in both the positive and negative directions. As a result, these jets have high positive and negative η values, making the overall sum of the $\cosh(\Delta\eta) - 1$ between all possible jet pairs large. In contrast, jets in the $6b$ resonant TRSM signal events have larger components in the transverse plane and are more homogeneously distributed, resulting in smaller sums of $\cosh(\Delta\eta) - 1$. This topology is also evident in the event shape sphericity shown in Figure 4.4(a). Furthermore, since the m_{HHH} is generally larger in $5b$ data than in $6b$ resonant TRSM signal events, their jets also tend to have larger p_T which are multiplied in the numerator of this variable. Hence $5b$ data events have larger values in the $\eta - m_{HHH}$ fraction than the $6b$ resonant TRSM signal, making it highly discriminating. For the NonresDNN, the m_H -radius in the (m_{H1}, m_{H2}, m_{H3}) volume is a variable dependent on jet pairing. As the m_{HHH} and its related variables between the SM HHH and background-like $5b$ data are similar, the more collimated jets from the Higgs boson decay reduce the jet pairing ambiguity in

Variable	Definition	NonresDNN	ResDNN	HeavyresDNN
m_H -radius	Euclidean distance between the event and the pairing centre (120, 115, 110) GeV in the (m_{H1}, m_{H2}, m_{H3}) volume.	✓		✓
$\cos \theta$	θ is the angle between the vector from the origin to the event's reconstructed mass of the Higgs boson candidates, and the vector from the origin to (120, 115, 110) GeV in the (m_{H1}, m_{H2}, m_{H3}) volume.		✓	
$\text{RMS}(m_{jj})$	Root-mean-squared (RMS) of the invariant mass of all possible jet pairs that can form a Higgs boson candidate.	✓		✓
$\text{RMS}(\Delta R_{jj})$	RMS of the angular separation between all possible jet pairs that can form a Higgs boson candidate.	✓	✓	✓
$\text{RMS}(\eta)$	RMS of the pseudo-rapidity of the Higgs boson candidates.	✓		✓
Skewness ΔA_{jj}	Skewness of $\cosh(\Delta \eta_{ik}) - \cos(\Delta \phi_{ik})$, where i, k are all possible jet pairs that can form a Higgs boson candidate.		✓	
Aplanarity _{6j}	The fraction of p_T from the 6 jets selected to reconstruct the 3 Higgs boson candidates lying outside the plane formed by the 2 highest p_T jets.	✓	✓	✓
Sphericity _{6j}	Isotropy of the momenta of the 6 jets selected to reconstruct the 3 Higgs boson candidates.		✓	
Transverse Sphericity _{6j}	Isotropy of the p_T of the 6 jets used to reconstruct the 3 Higgs boson candidates within the $x - y$ plane.	✓		
Sphericity	Isotropy of the momenta of all jets in the event.			✓

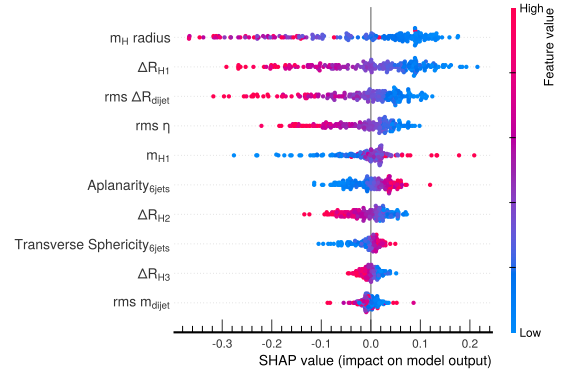
Table 4.6: Table summarising 10 out of the 16 input features selected for the ResDNN, NonresDNN and HeavyresDNN, and their definitions.

Variable	Definition	NonresDNN	ResDNN	HeavyresDNN
m_{H1}	Reconstructed mass of the leading p_T Higgs boson candidate.	✓		✓
$H_{T_{6j}}$	Scalar sum of the p_T of the 6 jets selected to reconstruct the 3 Higgs boson candidates.		✓	
$\eta - m_{HHH}$ fraction	$\frac{\sum_{i,j} 2p_T^i * p_T^j * (\cosh(\Delta\eta_{ij}) - 1)}{m_{HHH}^2}$ where i, j are all possible jet pairs that can form a Higgs boson candidate, and m_{HHH} is the reconstructed tri-Higgs invariant mass.		✓	
ΔR_{H1}	Angular separation between the jets paired to form the highest p_T Higgs boson candidate.	✓	✓	✓
ΔR_{H2}	Angular separation between the jets paired to form the second-highest p_T Higgs boson candidate.	✓	✓	✓
ΔR_{H3}	Angular separation between the jets paired to form the lowest p_T Higgs boson candidate.	✓	✓	✓

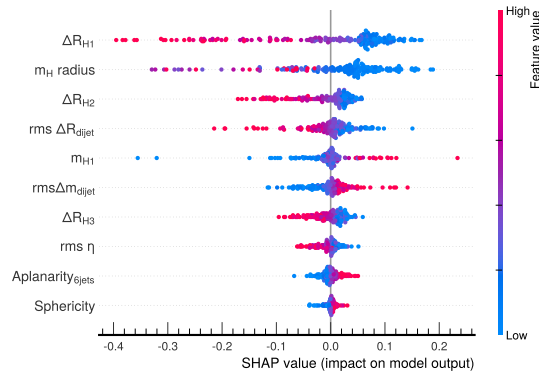
Table 4.7: Table summarising 6 out of the 16 input features selected for the ResDNN, NonresDNN and HeavyresDNN, and their definitions.



(a)



(b)



(c)

Figure 4.5: Shapley values of the input features selected for ResDNN 4.5(a), NonresDNN 4.5(b), and HeavyresDNN 4.5(c), ranked in descending order. The values are evaluated on the $5b$ data events as the background class, and the simulated $6b$ signal events used in the training of each DNN as the signal class.

the signal events. This in turn increases the pairing efficiency and forms Higgs boson candidates close to the mass centre in the (m_{H1}, m_{H2}, m_{H3}) volume. This is reflected in the smaller distance to this mass centre in the simulated signal compared to the $5b$ data. For the HeavyresDNN, ΔR_{H1} is most discriminating due to the highly collimated jets produced from the boosted Higgs bosons in the heavy-resonant signal events.

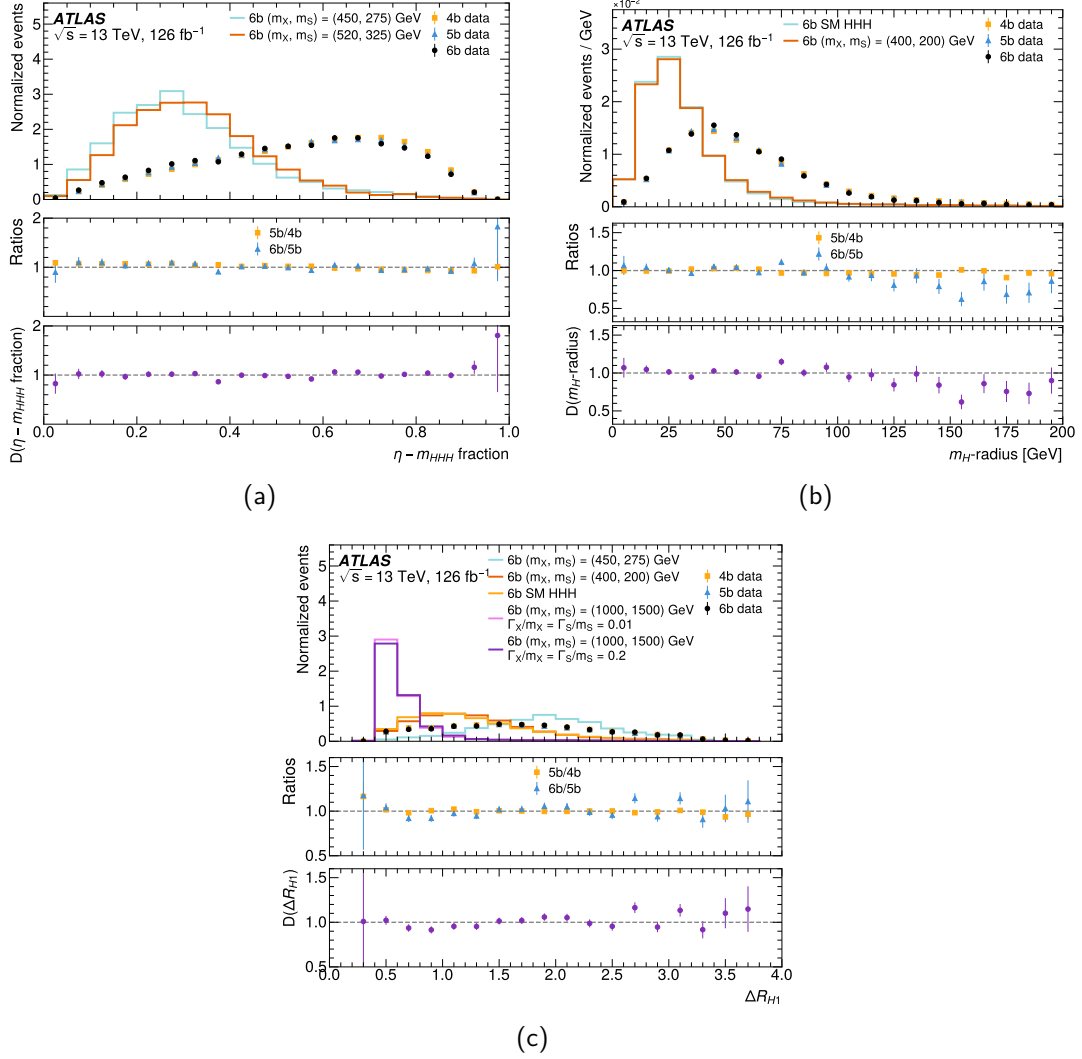


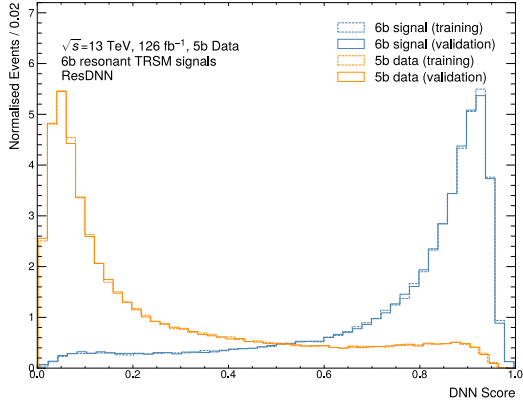
Figure 4.6: Distributions of the most impactful input feature of ResDNN (a), NonresDNN (b) and HeavyresDNN (c), evaluated on the relevant $6b$ signal and data in the $4b$, $5b$ and $6b$ regions. The middle panel shows the ratios of the distribution between $5b$ and $4b$ data, which are validated to have relatively small fluctuations. The bottom panel shows the double ratio D across the b -tag regions which is largely consistent with unity. The figures are published in [1].

4.4.3 Performance

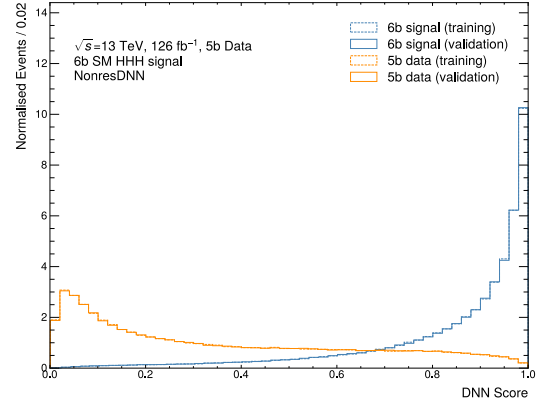
The output scores for $6b$ signal and $5b$ data events used in the training of the even and odd DNN models in the 2-fold validation are combined in each analysis interpretation, and are shown in Figure 4.7. The corresponding ROC curves for the models in each interpretation are shown in Figure 4.8. The AUC, which integrates over all TPR values in a ROC curve, is a good indicator of the overall performance of a model, whereas the shape of the ROC curve shows the FPR at a given TPR. The good agreement between both the output score distributions as well as the AUC values in the training and validation datasets indicate minimal overtraining in all models. All three types of models have an AUC over 0.9, which shows high efficiency in signal background separation. In particular, the HeavyresDNNs have the highest AUC, as the $6b$ heavy-resonant signal events have higher jet pairing efficiencies due to the more boosted Higgs bosons produced from the heavier scalar particles. This reduction in the ambiguity of jet pairing allows the kinematic variables of the Higgs bosons to be reconstructed more accurately, which is reflected in the DNN performance. The non-resonant interpretation also includes non-resonant HHH production via the TRSM, where the lighter scalar particle S is off-shell ($m_S < 250$ GeV) or where it is at the on- and off-shell border ($m_S = 250$ GeV). The former is referred to as the non-resonant TRSM signal, and the latter the border TRSM signal. Both the ResDNN and NonresDNN are applied to these signal events, and the one yielding the better performance is selected for the analysis. The ROC curve of these models evaluated on these $6b$ signal events are shown in Figure 4.9. The NonresDNN outperforms the ResDNN on both signal types, and is therefore selected.

4.5 Background Estimation

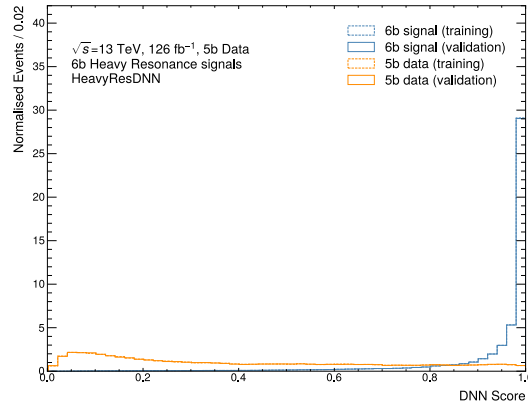
In order to identify the potential presence of a signal in data events inside the $6b$ Signal Region, it is important to model and understand the expected background events that are present. For a final state containing 6 b -jets, the dominant background process is QCD multijet production. Contributions from processes such as $t\bar{t}$ are small in this case, as the probability of mis-identifying 4 light-flavour jets in $t\bar{t}$ events as 4 b -jets by DL1d at the 77% working point is around 0.05%. This probability would increase with the presence of charm-jets (c -jets), as they are more easily mis-



(a)

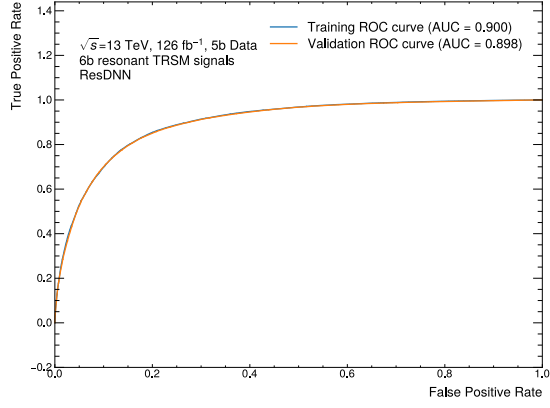


(b)

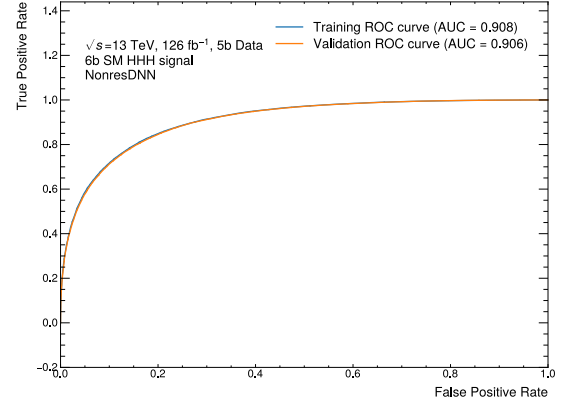


(c)

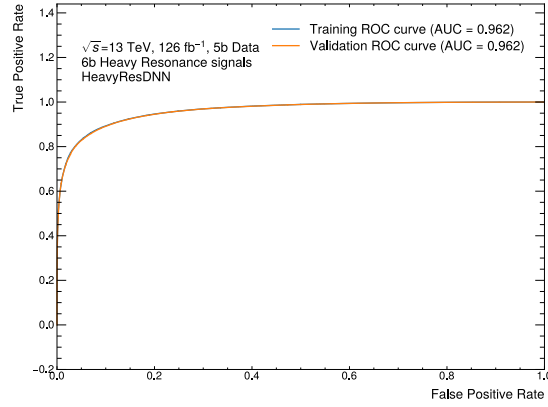
Figure 4.7: Output score distributions of the ResDNN (a), NonresDNN (b), and HeavyresDNN (c), where the outputs of the even and odd models in the 2-fold validations are combined.



(a)



(b)



(c)

Figure 4.8: The ROC curves of the ResDNN (a), NonresDNN (b), and HeavyresDNN (c), where the outputs of the even and odd models in the 2-fold validations are combined.

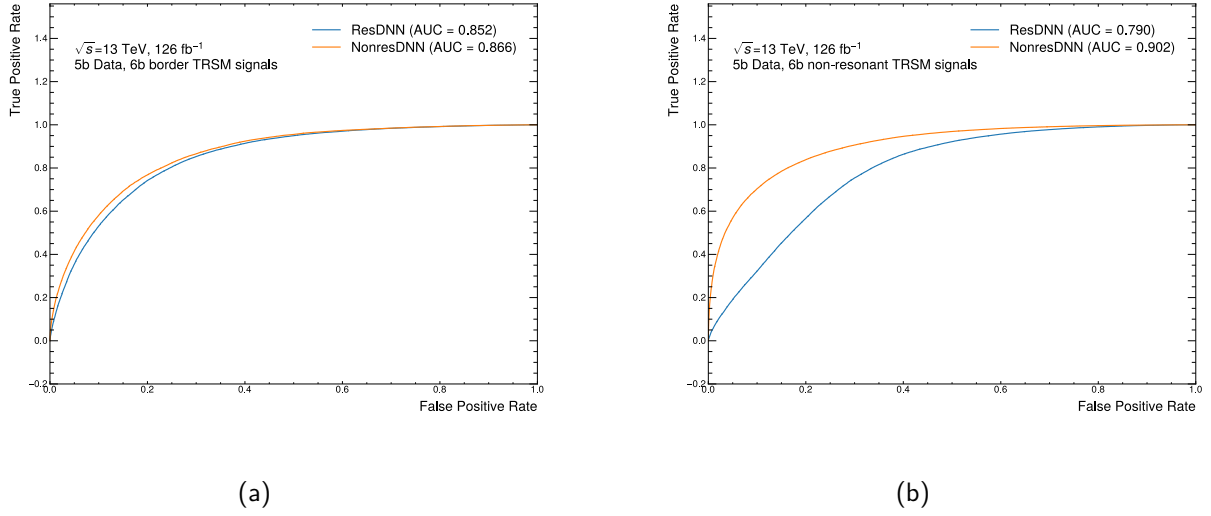


Figure 4.9: The ROC curves of the ResDNNs and NonresDNNs evaluated on the $6b$ border TRSM signal events where $m_S = 250$ GeV (a), and on the $6b$ non-resonant TRSM signal events where $m_S < 250$ GeV (b). The outputs of the even and odd models in the 2-fold validations are combined.

identified as b -jets. However, their production is rarer due to the higher mass of the charm-quark. Other background processes are tri-boson and di-Higgs productions, which have negligible contribution due to their small cross-sections. To estimate the DNN score distribution of the expected background in the $6b$ Signal Region, a fully data-driven method is used, outlined in Section 4.5.1. This method is motivated by the large uncertainties associated with simulated QCD multijet events, which make the modelling of their kinematic properties unreliable. The derived background estimate as well as the assumptions made during the derivation are validated in dedicated Control Regions, namely the $5b$ region and the Low-Score region of the DNNs, both of which are enriched by background events. The results and interpretations of these validations are described in Section 4.5.2. To account for the uncertainty arising from this estimate, systematic uncertainties are derived and outlined in Section 4.6.3.

4.5.1 Estimation Method

For the purposes of deriving and validating the background estimate, the recorded data events are divided into several regions by their b -jet multiplicity ($4b$, $5b$ and $6b$ regions) and DNN score (Low-Score and High-Score regions), illustrated in Figure 4.10. The detailed definitions of the scores regions for each DNN are summarised in Table 4.8. The DNN score distribution of the expected

background in the $6b$ region is derived from an extrapolation from the $4b$ to $5b$ then finally to the $6b$ region. This estimate is validated using events of all three b -jet multiplicities in the Low-Score regions, encased in the red box in Figure 4.10. This region is also used to derive the associated systematic uncertainties. The Low-Score regions are defined to only contain roughly 10% of the simulated $6b$ signal events in each interpretation, such that the $6b$ data in this region can be used before unblinding to validate the background estimate. The High-Score region is enriched in signal events, and combined with the Low-Score region, contains events used for statistical fitting and is referred to as the fitted region. Events in the region below the Low-Score region (Excluded region) are the least signal-like and fail to model the observed $6b$ data within one standard deviation of the statistical uncertainty. They are therefore excluded from the derivation of the background estimate and the statistical fitting.

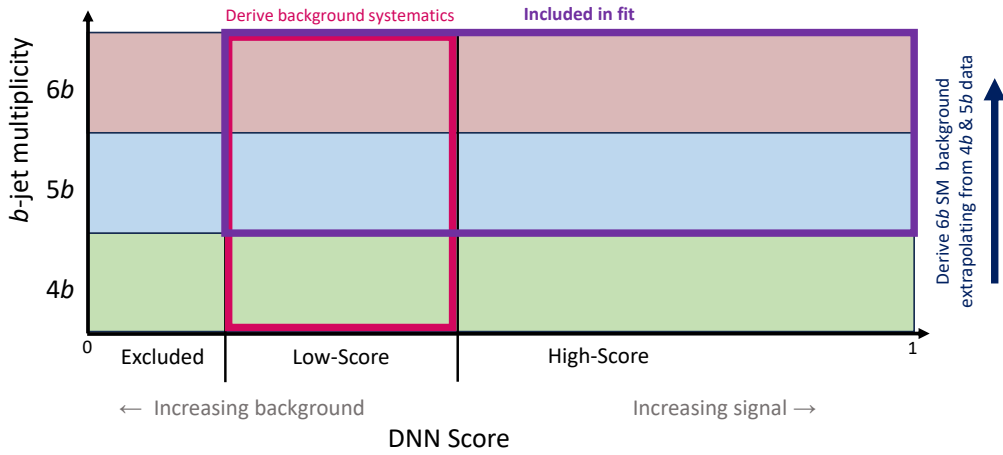


Figure 4.10: A schematic of the analysis regions defined using b -jet multiplicity and the DNN scores. The $6b$ region constitutes the Signal Region, whereas the $4b$ and $5b$ regions are used as Control Regions to derive the background estimate. The events are further divided by their DNN scores, where the Low-Score region contains more background-like events, and the High-Score contains more signal-like events. The Low-Score region is designed such that the $6b$ data can be unblinded to validate the background estimate and derive the systematic uncertainties associated with it. Events with very low DNN scores, which are ones that are the most background-like, are excluded from statistical fitting due to the poor modelling of the observed data in the $6b$ region. The figure is taken from [1].

The i^{th} bin in the DNN scores of the expected $6b$ background events is derived as:

$$N_i^{6b, \text{ estimated}} = N_i^{5b, \text{ fitted}} * \mu^{6b/5b} * \frac{\mu_i^{5b/4b}}{\mu^{5b/4b}} \quad (4.11)$$

DNN model	Excluded	Low-Score	High-score
ResDNN	$0 \leq \text{score} < 0.05$	$0.05 \leq \text{score} < 0.367$	$0.367 \leq \text{score} < 1$
NonresDNN	$0 \leq \text{score} < 0.1$	$0.1 \leq \text{score} < 0.552$	$0.552 \leq \text{score} < 1$
HeavyresDNN	$0 \leq \text{score} < 0.1$	$0.1 \leq \text{score} < 0.790$	$0.790 \leq \text{score} < 1$

Table 4.8: Definitions of the regions defined in the DNN scores of ResDNN, NonresDNN and HeavyresDNN. The boundaries between the Excluded and Low-Score regions are chosen such that good closure in the background-modelling is observed. The boundaries between the Low-Score and High-Score regions are chosen such that the contamination from $6b$ signals in each interpretation in the Low-Score region is around 10%. This amount of contamination is sufficiently low for the $6b$ data to be unblinded to validate the background estimate and derive its systematic uncertainties in the Low-Score region.

where $N_i^{5b, \text{fitted}}$ is the number of $5b$ events in the i^{th} bin of the DNN score determined during statistical fitting, $\mu_i^{5b/4b}$ is the scaling ratio between the observed $5b$ and $4b$ data yields in the i^{th} bin of the DNN score in the fitted region, and $\mu^{5b/4b}$ is the ratio of the overall yields in the same region. $\mu^{6b/5b}$ is the normalisation of the background estimate, determined in-situ during the statistical fitting, described in the next Section 4.8. With no signal present in the data, $\mu^{6b/5b}$ is simply N^{6b}/N^{5b} , and $N_i^{5b, \text{fitted}}$ is N_i^{5b} . However, if some signal is present, $\mu^{6b/5b} < N^{6b}/N^{5b}$, and $N_i^{5b, \text{fitted}} < N_i^{5b}$. The extrapolation of the DNN shape from the a low- b -tag to a high- b -tag region in the i^{th} bin is represented in the term $\frac{\mu_i^{5b/4b}}{\mu^{5b/4b}}$. This extrapolation is assumed to be similar to that from the $5b$ to the $6b$ region, and is therefore multiplied with the distribution of the $5b$ data, and normalised to the expected yield of the $6b$ background by $\mu^{6b/5b}$.

This method relies on two main assumptions:

1. The difference in the kinematic properties of the events is roughly the same each time the number of b -jets increases by one.
2. This extrapolation, validated using data in the Low-Score region of the DNNs where background events dominate, also holds in the High-Score region where signal events dominate.

These assumptions are validated using data events of all three b -jet multiplicities in the Low-Score regions of the DNNs, as well as only the low- b -tag regions ($4b$ and $5b$) in both the fitted region regions. The validity of the extrapolation across the b -tag regions is confirmed by the agreement between unity and the double ratio D , defined in Equation 4.10. The smoothness of the extrapolation

is indicated by the flat trend in D across the DNN score bins, shown in Figure 4.11. The details of these validation studies can be found in Section 4.5.2. After unblinding the entire $6b$ Signal Region, the DNN score distributions in the fitted regions of the $6b$ data events for the three analysis interpretations are shown in Figure 4.11. The double ratio D largely agree with unity within statistical uncertainties, with some deviations but no larger than 20%.

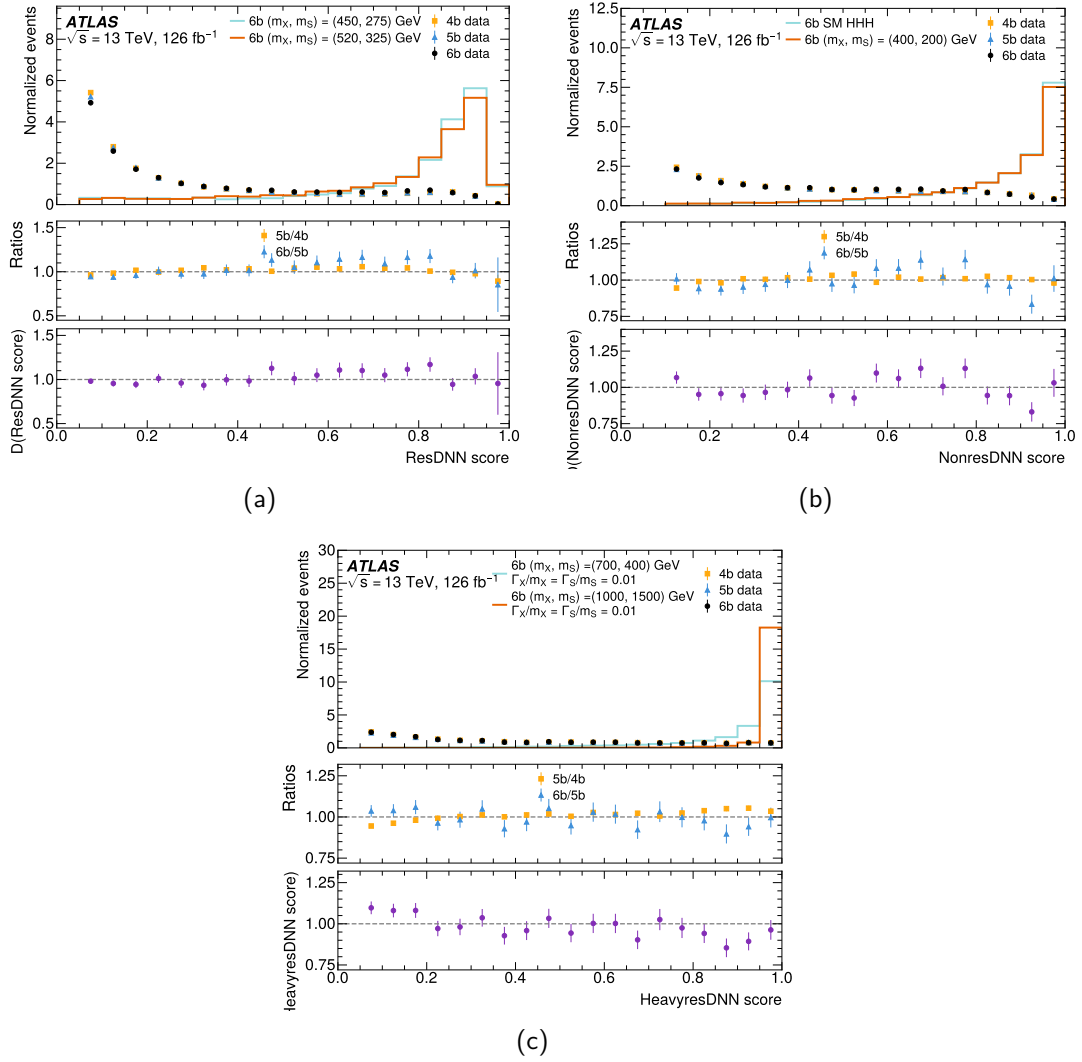


Figure 4.11: The output score distributions of the ResDNNs (a), NonresDNNs (b), and the HeavyresDNNs (c), for the representative signal samples of each interpretation, and for $4b$, $5b$ and $6b$ data in the corresponding fitted regions. The middle panel shows the ratios between high b -tag-region and low b -tag-region, where the relatively flat trends indicate smoothness of the extrapolation used to derive the background estimate. The bottom panel shows the double ratio D , where good agreement with unity is seen that indicates the validity of extrapolating across different b -tag regions. These figures are published in [1].

4.5.2 Validation

The assumptions made in deriving the background estimation are validated before unblinding the $6b$ Signal Region. The first assumption:

The difference in the kinematic properties of the events is roughly the same each time the number of b -jets increases by one.

is validated using the $4b$, $5b$ and $6b$ data events in the Low-Score regions of the DNNs. The validity is indicated again by the good agreement between the double ratio D and unity within statistical uncertainties in the fitted region, seen in Figure 4.12.

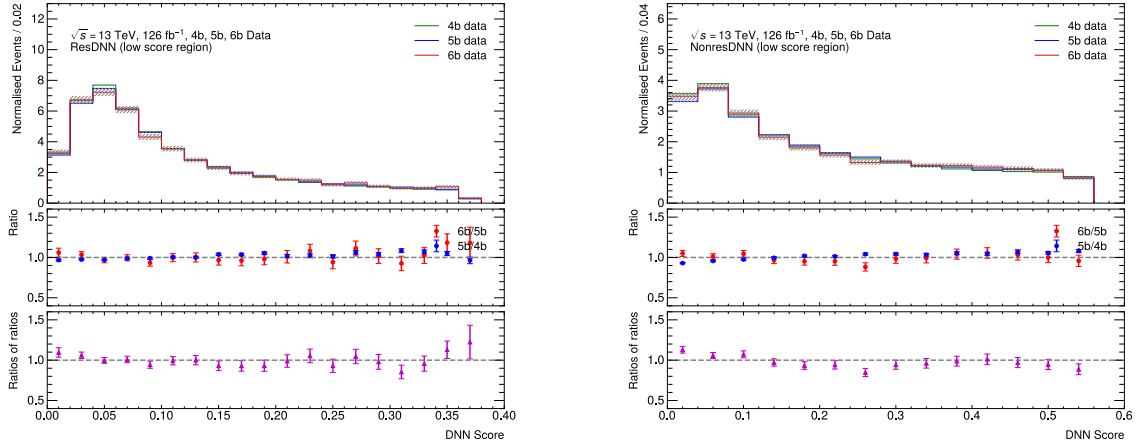
The second assumption:

The extrapolation, valid in the Low-Score region where background events dominate, also holds in the High-score region where signal events dominate.

is validated in the fitted region which contains both the Low-Score and High-score regions, with $4b$ and $5b$ data events. This is shown by the yellow points representing $5b/4b$ data ratios in Figure 4.11 in the middle panel, where they have a largely flat trend in the entire region.

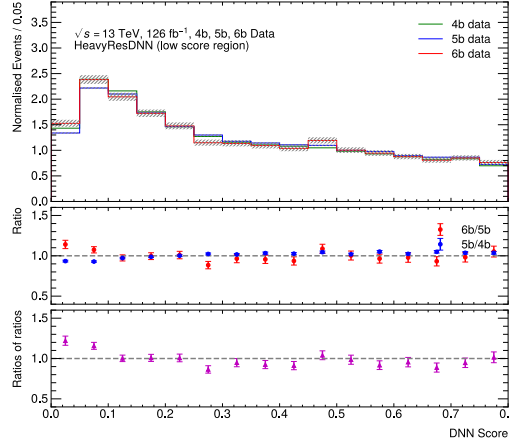
4.6 Systematic Uncertainties

Two types of uncertainties are considered in this analysis, statistical uncertainties and systematic uncertainties. For statistical uncertainties, the recorded data N follows a Poisson distribution, and has variance $\sqrt{N}$. The number of data in the relevant analysis categories, namely the $6b$, $5b$ and $4b$ data regions are summarised in Table 4.9. For systematic uncertainties, there are several sources: experimental uncertainties, uncertainties on the simulated signal Monte Carlo events from theory, and uncertainties associated with the data-driven background estimate. It is helpful to note that for this analysis, the statistical uncertainty on the observed data dominates due to the requirement of a large number of b -tagged jets.



(a)

(b)



(c)

Figure 4.12: The output score distributions of the ResDNNs (a), NonresDNNs (b), and the Heavyres-DNNs (c) for $4b$, $5b$ and $6b$ data in the corresponding Low-Score regions. The middle panel shows the ratios between high b -tag-region and low b -tag-region, where the relatively flat trends indicate smoothness of the extrapolation used to derive the background estimate. The bottom panel shows the double ratio D , where good agreement with unity is seen that indicates the validity of extrapolating across different b -tag regions.

Category	$4b$	$5b$	$6b$
Data yields (N)	1,925,520	140,606	8,244
Relative statistical uncertainty $\left(\frac{1}{\sqrt{N}}\right)$	0.00072	0.0027	0.011

Table 4.9: The observed data yields in each of the analysis categories after applying the event selections outlined in Section 4.3.3. The statistical uncertainty corresponding to each is $\sqrt{N}$ where N is the observed data yield.

4.6.1 Detector Uncertainties

Detector uncertainties in luminosity, jet reconstruction, jet vertex tagger (JVT) and b -tagging efficiencies, jet-level and b -jet level trigger efficiencies are considered. The derivation of uncertainties associated with trigger efficiencies is especially developed for this analysis, and is described in detail in Section 4.7. The remaining are summarised below:

- **Luminosity:** an uncertainty of 0.84% [92], using luminosity measurements obtained from a forward detector LUCID-2 [93].
- **Jet reconstruction:** jet energy scale (JES) that corrects the difference between the reconstructed jet energy and the real value, and jet energy resolution (JER) which is the precision at which jet energy is measured in the detector are considered by taking the values obtained from the calibration in [55], and applied to each jet.
- **JVT efficiency:** uncertainties associated with a neural network used to reject pile-up jets NNJVT are estimated using methods detailed in [52].
- **b -tagging efficiency:** the efficiencies of the b -tagging algorithm DL1d used to identify b -jets in simulated events and are corrected using methods detailed in [94] to match those in recorded data.
- **Trigger efficiencies:** uncertainties between 1 – 15% account for the difference in per-jet trigger efficiencies between simulation and recorded data events are applied to the simulated signal events. A flat uncertainty of 5% is also applied to account for the difference between b -tagging efficiencies at the trigger level and at the reconstruction level.

The relative impacts of various detector uncertainties on the expected upper limits placed on the cross-section of different types of simulated signal samples in the three analysis interpretations are summarised in Table 4.10.

Detector uncertainties	SM HHH	Non-resonant TRSM	Resonant TRSM	Heavy-resonant
Luminosity	<1%	<1%	<1%	<1%
Jet reconstruction	2.7%	2.3 - 6.5%	3.6 - 7.1%	1.0 - 6.3%
JVT & b -tagging	3.2%	2.8 - 5%	6.9 - 8.8%	1.5 - 5.6%
Trigger efficiencies	2.0%	1.8 - 3.5%	6 - 10%	1.4 - 4.2%

Table 4.10: The relative impacts of detector uncertainties on the expected upper limits placed on the cross-section of different types of simulated signal samples in the three analysis interpretations. The impacts are taken as the absolute percentage difference between the expected upper limits obtained from the profile likelihood fits with and without the systematic uncertainties in each case.

4.6.2 Theory Uncertainties

Uncertainties arise from the modelling inaccuracy of simulated signal events. The sources of these uncertainties are as follows:

- **Scales:** the renormalisation and factorisation scales are varied to estimate the uncertainties from missing higher-order terms in the perturbative expansion of the cross-section.
- **Strong coupling constant:** the uncertainty from the strong coupling constant α_s is estimated by using an alternative value.
- **Parton density functions:** the uncertainties arising from the choice of parton density functions (PDF) is taken as the standard deviation of 100 variations of the PDF set used in the simulation.

The relative impacts of the theory uncertainties on the expected upper limits placed on the cross-section of different types of simulated signal samples in the three analysis interpretations are summarised in Table 4.11.

Uncertainties	SM HHH	Non-resonant TRSM	Resonant TRSM	Heavy-resonant
Theory	1.5%	<1%	<1%	<1%

Table 4.11: The relative impacts of theory uncertainties on the expected upper limits placed for the cross-section of different types of simulated signal samples in the three analysis interpretations. The impacts are taken as the absolute percentage difference between the expected upper limits obtained from the profile likelihood fits with and without the theory uncertainties.

4.6.3 Background Estimation Uncertainties

The dominant systematic uncertainty in this analysis is associated with the shape of the background DNN score, estimated using the data-driven method described in Section 4.5.1. Uncertainties arise from the extrapolation from low b -tag into the high b -tag regions used in the estimation, and are derived using data in the Low-Score region. The non-closure in the DNN shape between different b -tag regions is broken down in the non-closure in individual kinematic variables used to train the DNNs. For a given kinematic variable, the non-closure is quantified by the deviations of their double ratio D from unity in the Low-Score region, for example the ones shown in Figure 4.4. This ensures the smoothness of the derived uncertainty, and captures the mis-modelling of background in a more physical way. The double ratio of a kinematic variable ν in the i^{th} DNN score bin of the Low-Score region is defined as:

$$D(i, \nu) = \left(\frac{N_{i,6b/5b}(\nu)}{\mu_{6b/5b}} \right)_{\text{Low-Score}} \div \left(\frac{N_{i,5b/4b}(\nu)}{\mu_{5b/4b}} \right)_{\text{Low-Score}} \quad (4.12)$$

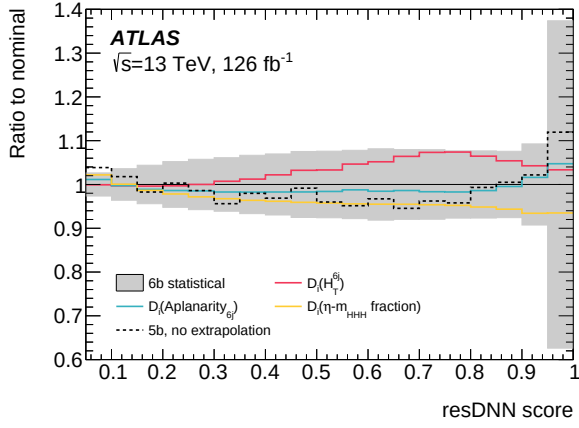
where the numerators are the ratios between high b -tag and low b -tag events in the i^{th} bin in the DNN Low-Score region as function of the kinematic variable ν . The denominators are the ratios between high b -tag and low b -tag events over the entire Low-Score region. To apply the shape uncertainties across all DNN score bins, where i now runs over all bins in the entire fitted region of the DNN score, the $D(i, \nu)$ is taken as a re-weighting factor for each event. Any deviation from $D(i, \nu) = 1$ captures the mis-modelling of background processes between different b -tag regions, and the mis-modellings in the Low-Score and High-Score regions are assumed to be similar. This assumption is validated using background-enriched $4b$ and $5b$ data in the fitted region, where the ratios between them show little fluctuations in Figure 4.11. The deviation from unity is symmetrised for each bin to form an up and a down multiplicative re-weighting factor for every event in each bin. The variations across the bins of the DNN score are fully correlated in the statistical fitting.

For the pair of DNNs used in each analysis interpretation, 10 kinematic variables are used to train each pair, which form 10 possible variations in the shape of the DNN score from varying each vari-

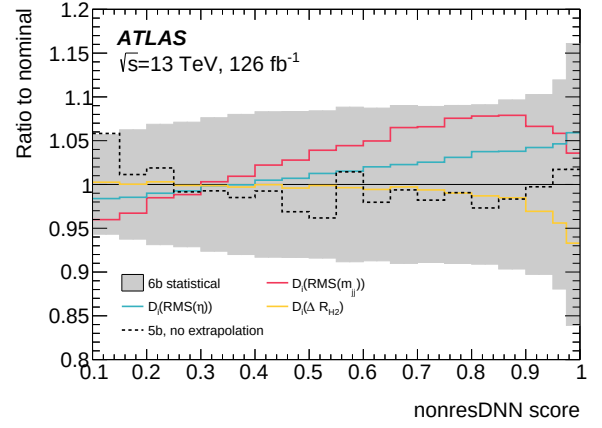
able. However, due to the correlations between these variables, the shape variations from some can be expressed as linear combinations of others. Therefore including all 10 shape variations would over-estimate the overall uncertainties on the DNN score. A pruning procedure is used to iteratively eliminate these redundant shape variations. It performs profile likelihood fits, iteratively adding the shape variation of a new variable ordered by the largest single-bin deviation in $D(i, \nu)$ iteratively, until χ^2 value, which characterises the goodness of fit, changes by less than 1 with the addition of a new variable. In other words, the variations in the DNN score shape are sufficiently described by the shape variations of the current set of variables in the iteration process. For each of the DNNs, a set of three variables are selected by the pruning procedure. The variations in the DNN score by varying each variable individually are shown in Appendix E. The ratios between the weighted and the nominal DNN score distributions by the shape variations of each variable are shown in Figure 4.13. There is some correlation between a feature importance, namely the Shapley value ranking, and the selected kinematic variables. For example, among the selected kinematic variables are the $\eta - m_{HHH}$, the most impactful input feature of the ResDNN, and ΔR_{H2} , one of the top three most impactful features for NonresDNN and HeavyresDNN. The weighted distributions by individual shape variations are all within the statistical uncertainty of the expected $6b$ background, despite the background shape uncertainty being the dominant systematic uncertainty. The relative impacts of the background shape uncertainty on the expected upper limits placed for the cross-section of different types of simulated signal samples in the three analysis interpretations are summarised in Table 4.12.

Uncertainties	SM HHH	Non-resonant TRSM	Resonant TRSM	Heavy-resonant
Background shape	16%	14 - 36%	18 - 30%	20 - 45%

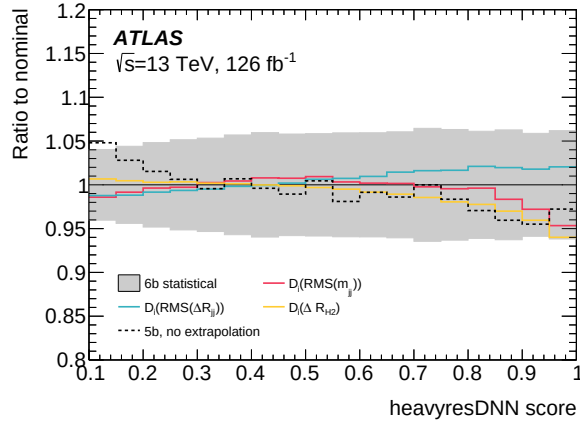
Table 4.12: The relative impacts of uncertainties associated with the background estimation on the expected upper limits placed on the cross-section of different types of simulated signal samples in the three analysis interpretations. The impacts are taken as the absolute percentage difference between the expected upper limits obtained from the profile likelihood fits with and without the background shape uncertainties included.



(a)



(b)



(c)

Figure 4.13: The DNN score distributions of the ResDNNs (a), NonresDNNs (b), and the HeavyresDNNs (c) in the fitted regions, weighted by the shape variations of the three kinematic variations selected by the pruning procedure. The weighted DNN score distributions are all within the statistical uncertainty of the expected $6b$ background, indicated in grey, and are on the same order in size as the statistical uncertainty of the observed $5b$ data. The figures are taken from [1].

4.7 Jet-level Trigger Calibration

The analysis uses both signal events generated from MC simulation, and recorded data events. Due to the differences in their kinematics, differences in the trigger efficiencies arise, especially in the low p_T region (the *turn-on* region) where the triggers are not fully efficient. As part of the event selection detailed in Section 4.3.3, events are required to have at least 4 jets with $p_T > 40$ GeV, and the Signal Region requires at least 6 b -tagged jets. Since genuine b -jets are more likely to have higher p_T than light-flavour jets, this implies that there are events that contain jets with p_T around 40 GeV or lower. For most hadronic triggers at ATLAS, the efficiencies are not 100% in this region, and are therefore more sensitive to kinematic differences that give rise to uncertainties. The b -jet triggers used to select events for this analysis from each data-taking year are documented in Section 4.3.1. They place requirements on 4 jets in every event at both the L1 and HLT levels. A dedicated method is developed to calibrate the trigger efficiencies to correct for these differences in the simulated signal events. The calibration provides a set of per-event *scale factors*, which are the ratios in the trigger efficiencies between MC and data events, parametrised by the p_T of the individual jets. The scale factors are multiplied with the event weight of each simulated signal event as a correction to its effective yield.

4.7.1 Simulated Samples and Event Selection

For this calibration, simulated semi- and di-leptonic $t\bar{t}$ MC and recorded data events that have similar kinematics but are orthogonal to the signal events simulated for the analysis are used. The top quarks in the simulated MC events have a near 100% decay branching fraction to a W boson and a b -quark, providing an abundance of b -jets for this calibration. The requirement of one or two leptons from the decay of the W bosons in the $t\bar{t}$ samples reduces the contributions from QCD multijet production and W/Z +jets backgrounds. In order to maximise high $t\bar{t}$ purity in both MC and data events, a reference trigger that is completely orthogonal to the b -jet triggers that are calibrated is used to select events by requiring a single muon and 2 additional jets. The reference triggers used for each data-taking year are summarised in Table 4.13.

Year	Reference trigger chain
2016	HLT_mu26_ivarmedium_2j35_boffperf_split
2017	HLT_mu26_ivarmedium_j110_gsc150_boffperf_split_j15_gsc35_boffperf_split
2018	HLT_mu26_ivarmedium_j110_gsc150_boffperf_split_j35_boffperf_split

Table 4.13: List of reference triggers used to select events for jet-level trigger SF derivation in MC and data. They all place requirements on a single muon and two additional jets at the HLT level. The conditions required at the HLT level follow the prefix "HLT_" and ones at the L1 level follow "L1_".

The detailed selection criteria placed on $t\bar{t}$ MC and data events for the calibration is summarised in the following:

- Pass the reference trigger for the corresponding data-taking year.
- Contain exactly one muon at the reconstruction level with $p_T > 30$ GeV and passing the FCTight isolation criteria.
- $E_T^{miss} > 20$ GeV, where E_T^{miss} is the missing transverse energy.
- $E_T^{miss} + m_{T,W} > 60$ GeV, where $m_{T,W}$ is the transverse mass of the W candidate defined as $m_{T,W} = \sqrt{2p_T^l E_T^{miss}(1 - \cos\phi_{l\nu})}$ from the leading p_T top-quark decay in $t\bar{t}$ events.
- Contain at least 4 jets with $p_T > 20$ GeV and $|\eta| < 2.5$, where at least 2 of them pass the 77% working point of the b -tagging algorithm DL1d.

The last criterion is closely aligned with the analysis event selection at the reconstructed level, and is imposed such that the events used for the calibration are kinematically similar to the events selected for the analysis.

4.7.2 Scale Factor Derivation Procedure

The procedure to derive the event-level scale factors follow the following steps. For the ease of terminology, objects defined at the trigger level are referred to as *online*, and ones defined at the reconstruction level are referred to as *offline*.

n^{th} jet ordering: Define the n^{th} jet as n^{th} offline jet ordered in descending E_T in both $t\bar{t}$ MC and data events.

Trigger matching: 4 of the ordered offline jets in every event are matched to an online object by geometric matching, where the pair of matched jets satisfy:

$$\Delta R = \sqrt{(\eta_{offline} - \eta_{online})^2 + (\phi_{offline} - \phi_{online})^2} < 0.4 \quad (4.13)$$

At the L1 level, the offline jet is matched to a L1 RoI, and at the HLT level to a jet reconstructed in the HLT (HLT jet). Since the analysis triggers place p_T requirements on 4 jets, only the first 4 *leading* jets that are trigger-matched are used in the calibration.

Jet-level trigger efficiency emulation: Emulate the jet-level L1 and HLT efficiencies $\varepsilon_{L1/HLT}$ as a function of p_T of the leading 4 trigger-matched offline jets. The L1 and HLT efficiencies can be interpreted as the following conditional probabilities:

$$\varepsilon_{L1}(n^{th}\text{jet}) = P(\text{matched to an L1 jet \& passed L1 threshold} \mid n^{th}\text{ jet } p_T) \quad (4.14)$$

$$\varepsilon_{HLT}(n^{th}\text{jet}) = P(\text{matched to a HLT jet \& passed the HLT threshold} \mid \text{event passed L1 thresholds \& } n^{th}\text{jet } p_T) \quad (4.15)$$

These are calculated for jets in both the $t\bar{t}$ MC samples and data events.

Event-level trigger efficiency: For each simulated HHH signal event, select the 4 leading jets that match to an online jet and multiply the corresponding jet-level efficiencies emulated for both $t\bar{t}$ MC and data from the previous step. Only 4 jets are required to pass the requirements at the trigger level. However, due to the ≥ 6 b -jets requirement in the Signal Region, an event could contain more than 4 trigger-matched offline jets. Therefore any 4 out of these jets could satisfy the trigger requirements, which gives rise to jet combinatorics. They are taken into account during the in-situ calculation of the event-level trigger efficiency of every HHH signal event. For example, for

an event with exactly 6 trigger-matched offline jets ($6j$), its event-level efficiency is calculated as

$$\varepsilon_{event}^{6j} = \prod_{m=1}^6 \varepsilon_m + \sum_{Set_{5j}} \left(\prod_{n \in Set_{5j}} \varepsilon_n \prod_{l \notin Set_{5j}} (1 - \varepsilon_l) \right) + \sum_{Set_{4j}} \left(\prod_{n \in Set_{4j}} \varepsilon_n \prod_{l \notin Set_{4j}} (1 - \varepsilon_l) \right) \quad (4.16)$$

where ε is the jet-level trigger efficiency parametrised in the p_T of the leading 4 trigger-matched offline jets derived from the previous steps. In the case where ≥ 4 offline jets are trigger-matched such as the example above, the efficiency of the 4^{th} jet is applied to the n^{th} jets where $n > 4$. Due to the relatively high jet p_T thresholds of the reference triggers, shown in Table 4.13, low statistics are observed in the trigger turn-on region for the leading trigger-matched offline jet, as only leading jets with high p_T are selected. As a solution, the jet-level trigger efficiencies for the sub-leading trigger-matched jets are used for the leading jets. The first term in Equation 4.16 represents the efficiency where all $6j$ pass the trigger requirements, the second and the third terms represent when $5j$ and $4j$ out of the $6j$ pass the trigger requirements. The event-level efficiencies at the L1 and HLT levels are calculated separately.

Event-level scale factor: Finally, the event-level scale factor for each simulated HHH signal event is calculated by taking the ratios in the event-level trigger efficiencies calculated from $t\bar{t}$ MC and data events. This is multiplied with the signal event's event weight to correct for its representative yield.

4.7.3 Calibration Results

First the matching efficiencies between offline jets and L1 Rols, and HLT jets are defined as the following:

$$\varepsilon_{L1}^{matching} = P(\text{an offline jet trigger-matched to a L1 Rol}) \quad (4.17)$$

$$\varepsilon_{HLT}^{matching} = P(\text{an offline jet trigger-matched to a HLT jet} \mid \text{event passed L1 thresholds}) \quad (4.18)$$

The L1 and HLT matching efficiencies, for $t\bar{t}$ MC and data events are shown for the 2018 data-taking year in Figures 4.14 and 4.15 as representative results. The low statistics of the leading (1st) L1 matched offline jet is evident as a result of the higher jet p_T thresholds, but sufficient statistics

are observed for the remaining jets. The matching efficiencies across the n^{th} jets largely overlap, indicating little bias introduced by the reference trigger and event selections used for the calibration. The HLT matching efficiencies are near 100% for most of the p_T region as the events that pass L1 trigger thresholds contain jets with higher p_T , which are more likely to match to a HLT jet.

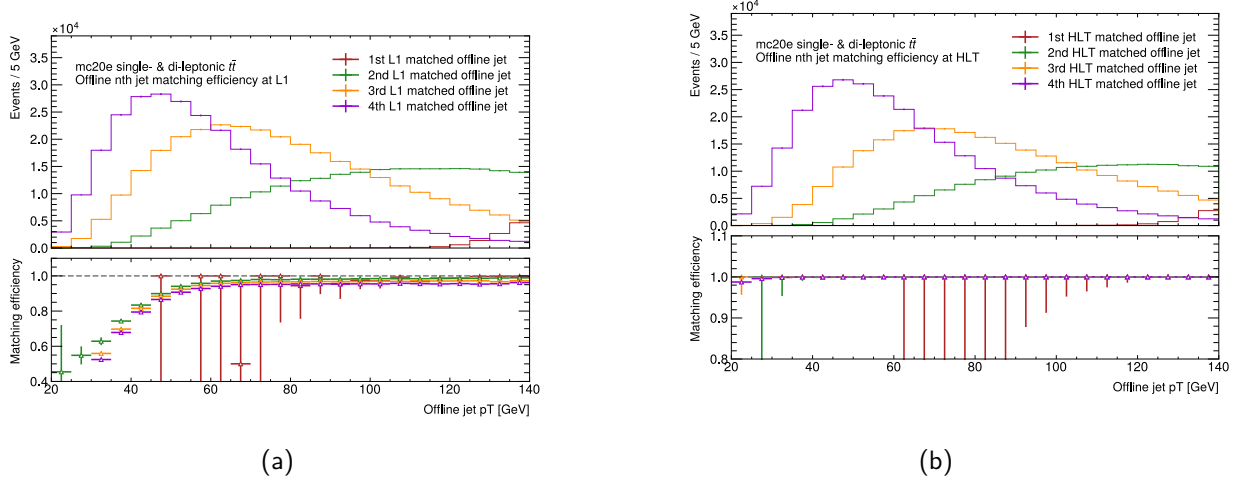


Figure 4.14: The matching efficiencies of the 4 leading offline jets to L1 Rols (a), and to HLT jets (b) in $t\bar{t}$ MC samples for the 2018 data-taking year condition (mc20e).

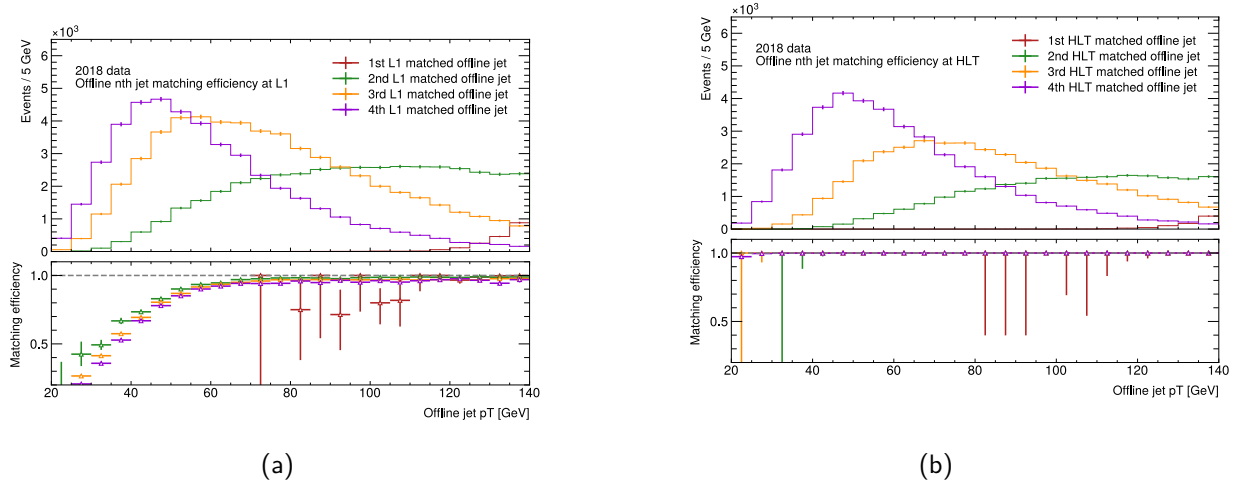


Figure 4.15: The matching efficiencies of the 4 leading offline jets to L1 Rols (a), and to HLT jets (b) in data for the 2018 data-taking year condition.

The jet-level trigger efficiencies for the n^{th} jets used to derive the event-level trigger efficiencies are shown in Figures 4.16 and 4.17 for the L1 and HLT respectively. Efficiencies for only the 2^{nd} , 3^{rd} and 4^{th} jets are shown as the leading jets are not used due to low statistics.

The event-level trigger scale factors at the L1 and HLT level are shown in Figure 4.18 separately for

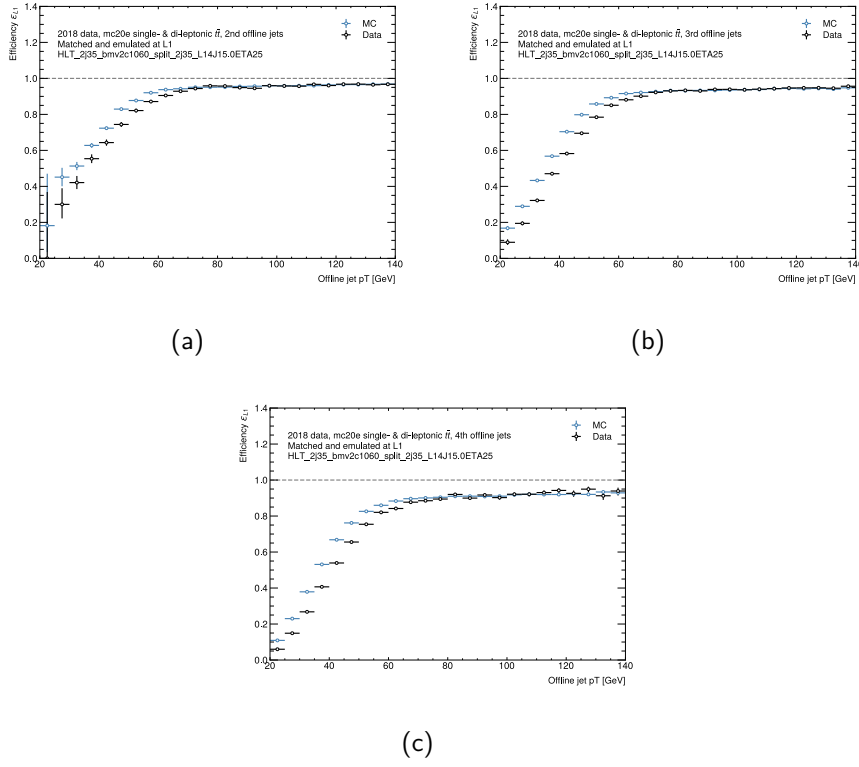


Figure 4.16: The jet-level L1 trigger efficiencies of the 2^{nd} , 3^{rd} and 4^{th} jets in the $t\bar{t}$ MC and data events for the 2018 data-taking year condition (mc20e).

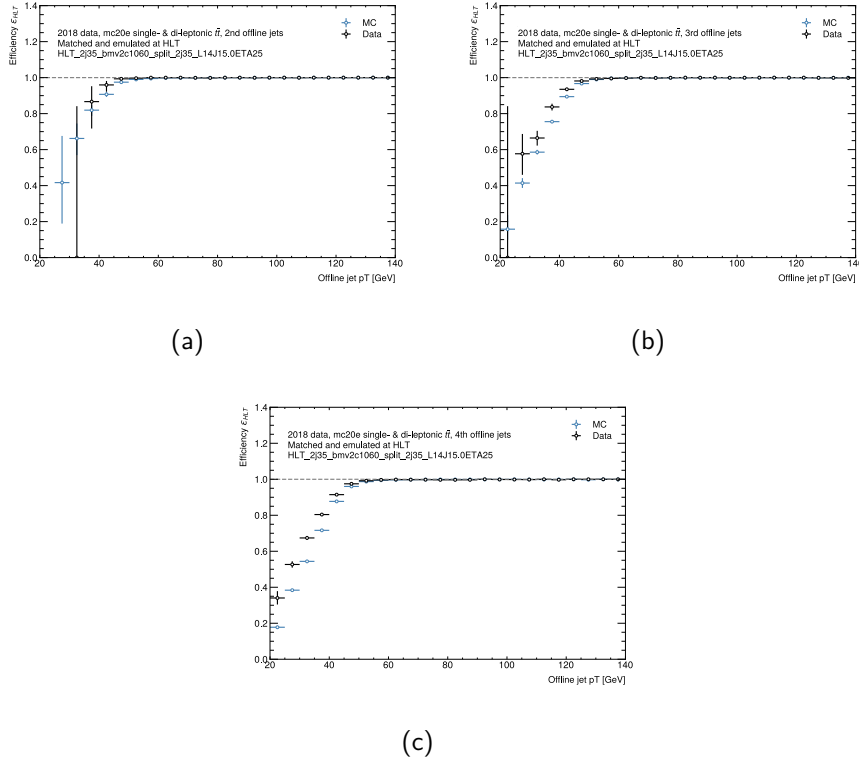


Figure 4.17: The jet-level HLT trigger efficiencies of the 2^{nd} , 3^{rd} and 4^{th} jets in the $t\bar{t}$ MC and data events for the 2018 data-taking year condition (mc20e).

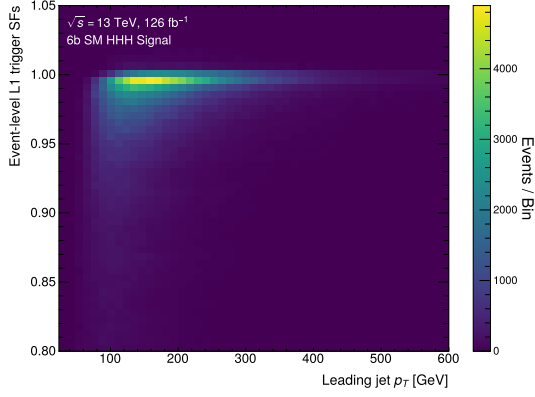
the representative signal samples in each analysis interpretation. The deviations of the scale factors from 1 are larger at L1 compared to HLT due to the lower trigger efficiencies at L1. The deviations are also larger for signal samples with jets that have lower p_T , for example for the SM HHH and the resonant TRSM signals. Again, this is due to lower trigger efficiencies in the low jet p_T region. As a result, the effective yields in these signal samples are reduced mainly by the L1 trigger scale factors.

4.7.4 Systematic Uncertainties

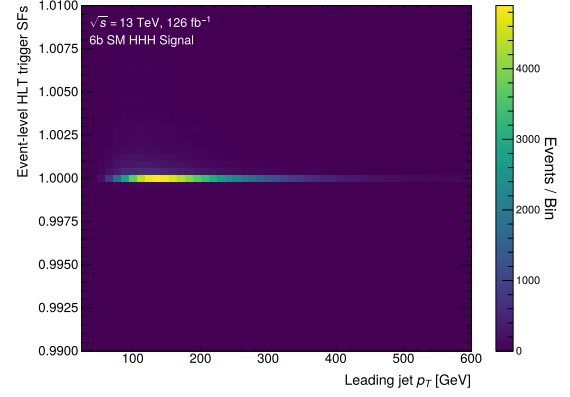
Systematic uncertainties on the scale factors can arise from the choice of matrix element and parton showering used to simulate $t\bar{t}$ MC samples. The nominal $t\bar{t}$ MC samples are generated with the Powheg generator, and the parton showering is handled with the A14 tuned Pythia8 (PhyPy8EG) [86]. Two alternative samples are used to derive the systematic uncertainties from the $t\bar{t}$ modelling, one with an alternative matrix element NLO matching scale (pThard), and the other with an alternative parton shower using Herwig7 [95]. The differences in the jet-level trigger efficiencies between the nominal and the alternative samples are combined as quadrature sums, and propagated as binomial errors as the overall systematic uncertainties on the event-level scale factors. These uncertainties are calculated in-situ together with the event-level trigger scale factors, and used as the nominal values of a nuisance parameter in the statistical fitting. The jet-level L1 and HLT efficiencies of the nominal and alternative $t\bar{t}$ samples for the 2018 data-taking year are shown in Figures 4.19 and 4.20. The systematic uncertainties in populous bins are at the percentage level in both L1 and HLT efficiencies.

4.7.5 b -jet level Trigger Uncertainty

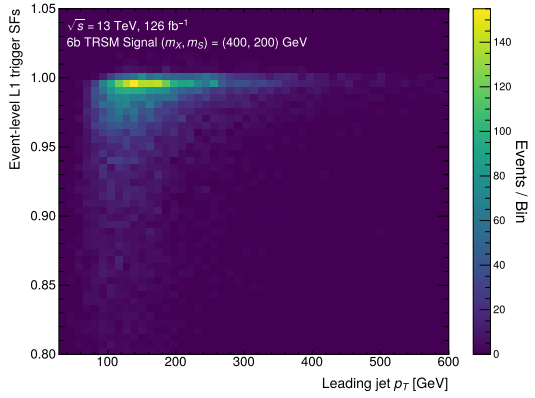
Differences in trigger efficiencies can also arise from the difference in the b -tagging efficiency at the trigger level (online b -tagging) with respect to reconstruction level (offline b -tagging), between simulated MC and data events. This calibration is historically performed centrally for previous b -tagging algorithms in the trigger, for example in [96]. The efficiencies, calculated as the probability of a jet passing a working point of online b -tagging, given that it has passed a working point of



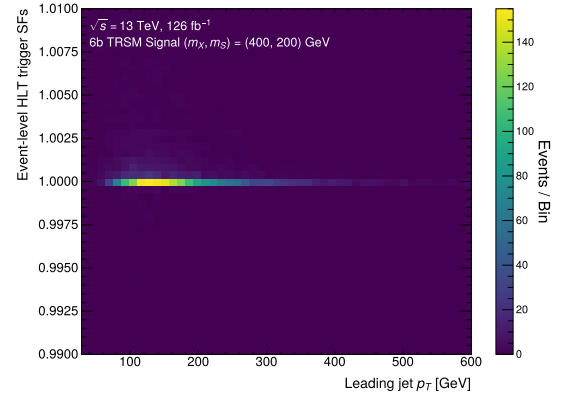
(a)



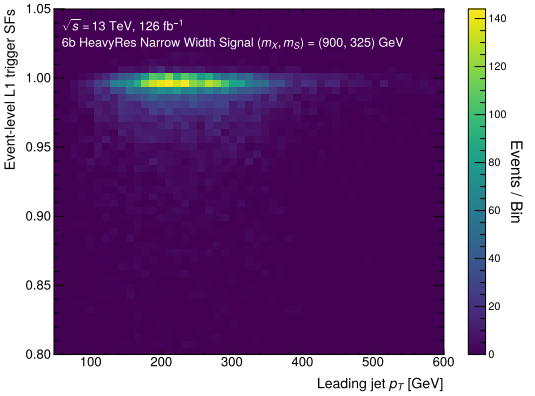
(b)



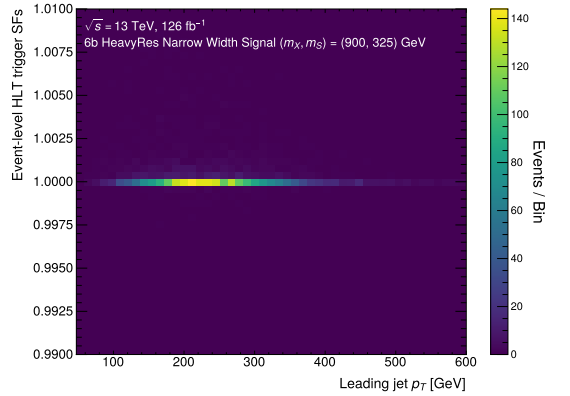
(c)



(d)



(e)



(f)

Figure 4.18: The event-level L1 and HLT trigger scale factors as functions of the leading offline jet p_T for the $6b$ SM HHH signal ((a), (b)), a representative sample of the resonant TRSM signals ((c), (d)) and one for the heavy-resonant signals ((e), (f)). The scale factors are used to correct the yields in the sample from discrepancies in trigger efficiencies between data and simulated MC events.

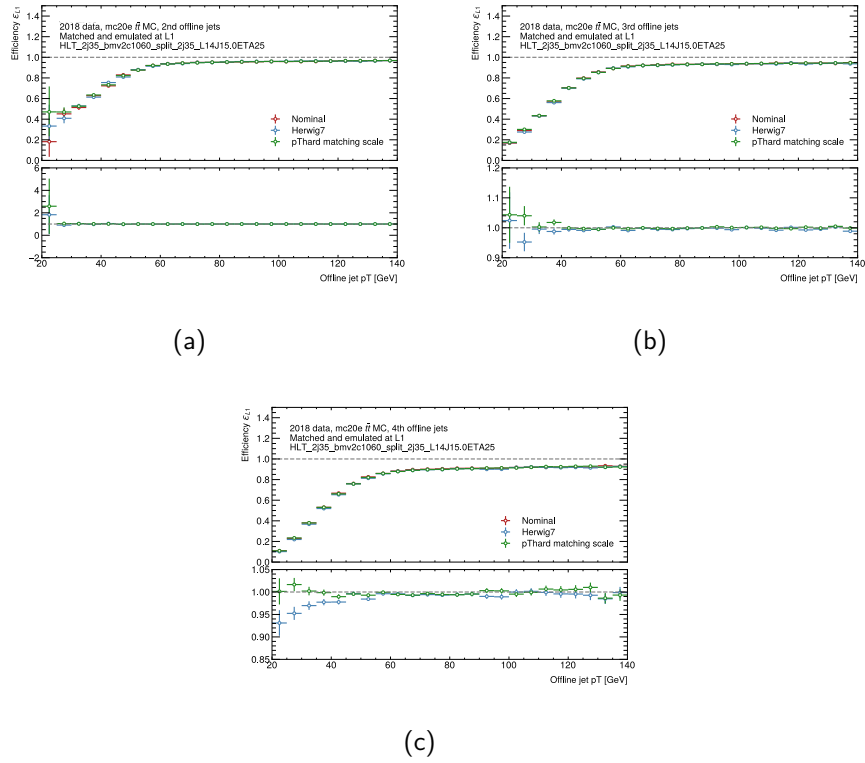


Figure 4.19: The jet-level L1 trigger efficiencies of the 2^{nd} , 3^{rd} and 4^{th} jets in the nominal and alternative $t\bar{t}$ samples as functions of the offline jet p_T for the 2018 data-taking year condition (mc20e). The error bars show the statistical errors in each bin.

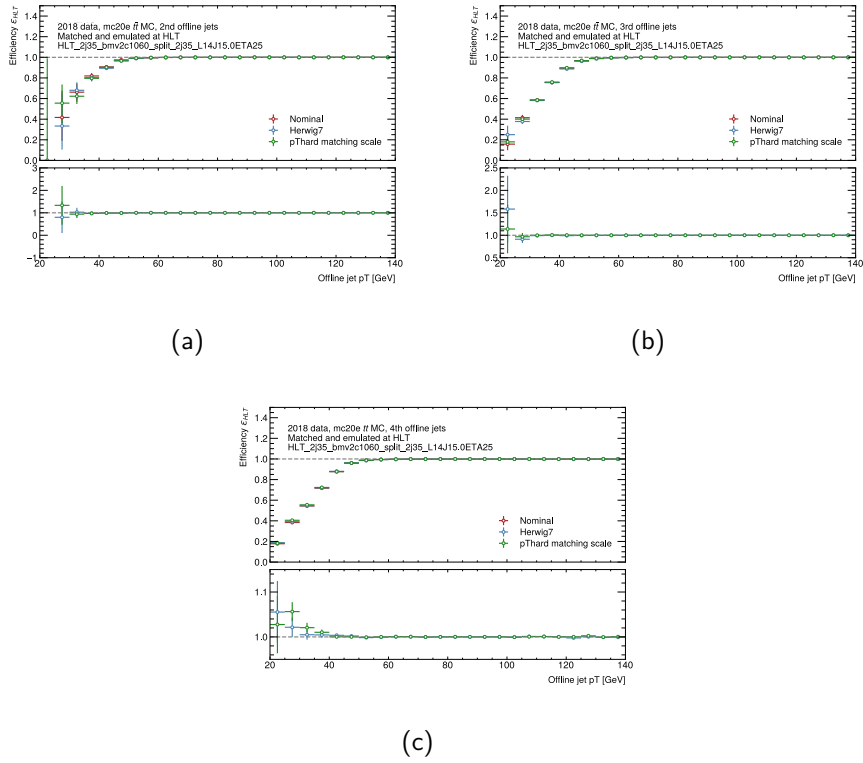


Figure 4.20: The jet-level HLT trigger efficiencies of the 2nd, 3rd and 4th jets in the nominal and alternative $t\bar{t}$ samples as functions of the offline jet p_T for the 2018 data-taking year condition (mc20e). The error bars show the statistical errors in each bin.

offline b -tagging, are measured in both data and MC, and their ratios as functions of the jet p_T are taken as the scale factors at the b -jet level. However, the lack of trigger matching information from the b -jet triggers meant the calculation of these scale factors was not possible. In addition, the uncertainties from the b -jet triggers would have an insignificant effect in this analysis for the following reasons:

1. The background estimate is fully data-driven, therefore only the yields of the simulated signal events need to be corrected.
2. The Signal Region requires at least 6 b -tagged jets at the reconstruction level, whereas only 2 or 3 b -tagged jets are required at the trigger level. The probability of an event with 2 or 3 jets passing the online b -tagging requirement given at least 6 jets already passing the offline b -tagging is high. This conditional probability is estimated to be around 0.99 for all $6b$ resonant TRSM signal samples.

Due to the high online b -tagging efficiency given the offline b -tagging requirements, the size of its uncertainty would be smaller than the inefficiency, which is at most 1% for $6b$ resonant TRSM signal events. This inefficiency would be even smaller for the other types of signal events in the analysis, as jets in the resonant TRSM signal events have the lowest p_T , and therefore the lowest b -tagging efficiency. Combining the above reasons, a fixed 5% is assigned as the trigger uncertainty at the b -jet level.

4.8 Statistical Analysis

The final step of the analysis is to compare recorded data to theory in order to search for new physics, or draw conclusions about the underlying parameters of the theory. A likelihood-based statistical test [97] is performed, where a test statistics is built based on the likelihood function, and the compatibility between the observed and the expected data is quantified by the p -value, detailed in Section 4.8.1. The definition of the test statistics and the CLs method used to set upper limits on the production cross-section of the SM and BSM HHH production cross-sections is outlined in Sections 4.8.2 and 4.8.3. Finally the statistical setups for the setting the upper limits on the

production cross-sections and the simultaneous scanning of κ_3 and κ_4 are detailed in Section 4.8.4.

4.8.1 The Likelihood Function and p -value

For a random variable x with a probability density function (PDF) $f(x, \boldsymbol{\mu})$, where $\boldsymbol{\mu}$ are some parameters that parametrise it, the likelihood function after n repeated measurements can be expressed as the joint PDF of the observed data x_i :

$$\mathcal{L} = \prod_{i=1}^n f(x_i; \boldsymbol{\mu}) \quad (4.19)$$

assuming that the measurements are independent and fixed. The best estimators of the parameters $\hat{\mu}_i$ maximise the likelihood function, and are given by the solutions to $\frac{\partial \ln \mathcal{L}}{\partial \mu_i}$, where the derivative of the logarithm is solved for ease of calculation but yielding the same optimal estimators. As the explicit evaluation of a likelihood function is often intractable in particle physics, data events are often presented in histograms of some summary statistics, otherwise known as the main discriminant. For example, a natural choice for the search of a hypothetical particle is its mass. In the case of the $HHH \rightarrow 6b$ analysis, the DNN score of each interpretation is chosen as the main discriminant to maximise signal sensitivity. The likelihood function can therefore be expressed as the product of the joint probability function of events in each bin of the summary statistics histogram:

$$\mathcal{L} = \prod_{b \in \text{bins}} f_{\text{Poisson}}(n_b | \lambda_b(\boldsymbol{\mu}, \boldsymbol{\theta})) \times \prod_{\chi \in \mathcal{X}} g_{\chi}(y_{\chi} | \chi) \quad (4.20)$$

where n_b is the observed number of events in bin b , and $\lambda_b(\boldsymbol{\mu}, \boldsymbol{\theta})$ is the expected number of events given the parameters of interest (POIs) $\boldsymbol{\mu}$ and nuisance parameters (NPs) $\boldsymbol{\theta}$ which describe the uncertainties of this expected number of events. The joint probability functions are Poisson distributed, as it naturally describes events occurring at a fixed rate which is common in most particle physics experiments. The second term in the likelihood function describes a set of NPs χ which form a subset of $\boldsymbol{\theta}$ that are constrained by some auxiliary measurements y . These auxiliary measurements are made independently of data considered in the target analysis. They provide additional information in order to reduce the impact of these uncertainties. For statistical uncertainty, g_{χ} would be a Poisson distribution, and for other systematic uncertainties, g_{χ} is often a Gaussian distribution.

The inclusion of NPs here capture various sources of uncertainties that could yield an inaccurate prediction of the expected number of events λ_b . They are often applied as multiplicative or additive factors accounting for the up and down variations of the expected number of events in each bin.

To test a certain hypothesis, the compatibility of data with a certain hypothesis is quantified by the p -value:

$$p = \int_{t_{obs}}^{\infty} f(t|H) dt \quad (4.21)$$

where t is the test statistics for the observed data, and $f(Q|H)$ is its distribution given a hypothesis H . The p -value represents the probability of observing a value of Q that is more or as extreme as Q_{obs} . The confidence level, $(1 - \alpha)$, where α is a chosen value, is the probability that the true value of the parameter lies within the confidence interval. The hypothesis would be rejected if $p \leq \alpha$ at a confidence level of $(1 - \alpha)$. The p -value can be expressed as a significance Z measured in standard deviation (σ), such for a variable that follows a Gaussian distribution Φ , a value that is Z standard deviations above the mean has an upper-tail probability equal to the p -value.

$$\begin{aligned} p &= 1 - \Phi(Z) \\ Z &= \Phi^{-1}(1 - p) \end{aligned} \quad (4.22)$$

In particle physics, to reject a background-only hypothesis, a significance of at least $Z = 5$ is the convention to constitute a discovery. To exclude a signal hypothesis, a confidence level of 95%, equivalent to a p -value of 0.05 and $Z = 1.64$ is often used.

4.8.2 The Test Statistics

According to the Neyman-Pearson Lemma [98], given a significance level α , the most powerful test statistic is the likelihood ratio [97], expressed as

$$\Lambda = \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}(\mu))}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \quad (4.23)$$

where the denominator is the unconditional maximum likelihood, where $\hat{\mu}$ and $\hat{\theta}$ are the maximum-likelihood (ML) estimators that yield a hypothesis most consistent with the observed data. The numerator is the conditional maximum likelihood, where $\hat{\theta}(\mu)$ are ML estimators that maximise $\mathcal{L}$ given a fixed μ . The NPs θ are a function of the POI μ , which broaden the profile likelihood as a result of the presence of systematic uncertainties on μ . Commonly, $t = -2 \ln \Lambda(\mu)$ is used instead as the test statistic, known as the negative log-likelihood ratio. In the $HHH \rightarrow 6b$ search, the POI of the statistical test is the signal strength $\mu = \mu_{sig}$ to test for the presence of a signal. Since the signal strength is defined to be positive, the test statistic becomes

$$q_\mu = \begin{cases} -2 \ln \Lambda(0) & 0 \leq \hat{\mu} \leq \mu \\ 0 & \text{otherwise} \end{cases} \quad (4.24)$$

referred to as the one-sided profile log likelihood ratio. It compares a specific hypothesis $H(\mu)$ with one that maximises the global likelihood, and q_μ approaches 0 when the compatibility between data and the hypothesis increases.

4.8.3 The CLs Method

The CLs procedure is often used in particle physics to set upper limits for production cross-sections. It is a conservative method in the sense that, the upper limit being greater than the true value has a probability that is equal or greater than the chosen confidence level. This conservative approach is beneficial when the distribution of the test statistic of the background-only hypothesis H_0 and the signal hypothesis H_1 overlap significantly, which is often the case when searching for rare or BSM physics processes. The CLs method is therefore commonly adopted to avoid false discovery of new physics at the LHC. The signal hypothesis is excluded if

$$\begin{aligned} \text{CL}_s &= \frac{\text{CL}_{s+b}}{\text{CL}_b} = \frac{\int_{-\infty}^{t_{obs}} f(t_\mu | H_1) dt_\mu}{\int_{-\infty}^{t_{obs}} f(t_\mu | H_0) dt_\mu} \\ &= \frac{p_{s+b}}{1 - p_b} < \alpha \end{aligned} \quad (4.25)$$

Here, p_{s+b} and p_b are the p -values given H_1 and H_0 respectively, and α is the confidence interval. In order to place upper limits at a given confidence interval on, say the signal strength of a new physics signal, p_s can be scanned over a range of values for μ_{sig} to find the value corresponding to $CL_s = \alpha$. Therefore, hypotheses with μ_{sig} greater than this value represent physics processes that the experiment is not sensitive to presently, and can be excluded at this confidence level.

It is common before the start of an analysis to investigate the *expected* sensitivity to the targeted new physics processes. In the case of $HHH \rightarrow 6b$, expected upper limits on the signal cross-section is given before the analysis is unblinded¹. They test the expected upper limits on the various signal cross-sections given different values using Asimov datasets². An Asimov dataset is a single representative dataset, which in this case consists of only background-like events for the estimation of signal cross-section upper limits. The signal cross-section upper limits obtained from the unblinded dataset are referred to as the *observed* upper limits.

4.8.4 Fit Setup

The likelihood function of the observed number of events in data is

$$\begin{aligned}
\mathcal{L}(\vec{N}^{5b,6b}, \vec{S}^{5b,6b} | \vec{\theta}, \vec{\phi}, \vec{\gamma}^{5b,6b}, \mu_s, \mu_{NF}) = & \prod_{i \in \text{bins}} \left[\text{Pois} \left(N_i^{6b} | \mu_{NF} f_i(\vec{\theta}) B_i + \mu_s w_i^{6b}(\vec{\phi}) \gamma_i^{6b} S_i^{6b} \right) \right. \\
& \times \text{Pois} \left(N_i^{5b} | B_i + \mu_s w_i^{5b}(\vec{\phi}) \gamma_i^{5b} S_i^{5b} \right) \Big] \\
& \times \prod_{m \in \text{bkg. sys.}} \frac{1}{\sqrt{2\pi}} e^{-\theta_m^2/2} \\
& \times \prod_{k \in \text{sig. sys.}} \frac{1}{\sqrt{2\pi}} e^{-\phi_k^2/2} \\
& \times \prod_{i \in \text{bins}} \left[P_\Gamma \left(\nu_i^{5b} | \gamma_i^{5b} \nu_i^{5b} \right) P_\Gamma \left(\nu_i^{6b} | \gamma_i^{6b} \nu_i^{6b} \right) \right].
\end{aligned} \tag{4.26}$$

¹In particle physics, the most signal-like observed data is often not used, or 'blinded', during the optimisation of the analysis strategy in order to reduce bias.

²The name of the dataset is inspired by the science fiction short story *Franchise* by Issac Asimov, where a single representative voter is chosen to represent the entire electorate.

The numbers of observed data in the $5b$ and $6b$ regions are denoted $N^{\vec{5b,6b}}$, and the numbers of expected signal events are $\vec{S}^{5b,6b}$. The expected number of background events in each bin are B_i , which are obtained from unconstrained bin-by-bin fits from the Control Regions. To place upper limits on the production cross-section, the POI of the fit is the signal strength denoted μ_s . To place limits on the Higgs self-coupling modifiers κ_3 and κ_4 , the POIs become the two modifiers in the likelihood expression.

In the first product over all bins in the DNN score fitted region, the Poisson terms describe the distributions of the observed data in each bin N_i^{5b} , N_i^{6b} in the $5b$ Control Region and the $6b$ Signal Region respectively. The non-negligible amount of signal contamination in the $5b$ Control Region is taken into account by the term S_i^{5b} . The $\vec{\gamma}^{5b,6b}$ denotes the NPs accounting for the statistical uncertainty of the MC samples in the $5b$ and $6b$ regions. The NPs describing the systematic uncertainties in the simulated signal samples used as the template for the signal hypothesis are denoted $\vec{\phi}$, and are fully correlated between the $5b$ and $6b$ regions. The NPs describing the systematic uncertainties associated with the background estimate are denoted $\vec{\theta}$. The NPs affect the yields in each bin with $w_i(\vec{\phi})$ and $f_i(\vec{\theta})$, where $w_i(\vec{\phi})$ only affects terms associated with the signal, and $f_i(\vec{\theta})$ only affects those concerning the background estimate of the $6b$ region. The nominal value of $w_i(\vec{\phi})$ is 1, which captures the impacts of the theoretical and experimental uncertainties on the shape and normalisation of the signal. The nominal value of $f_i(\vec{\theta})$ is the bin-by-bin ratio of the DNN score between $5b$ and $4b$ data, $((N^{5b}/N^{4b})_i)/(N^{5b}/N^{4b})$, corresponding to the term in the derivation of the background estimate in Equation 4.11. The $f_i(\vec{\theta})$ describes the extrapolation of the background DNN score shape from the $5b$ to $6b$ region. A normalisation term μ_{NF} , responsible for scaling the overall yields of the background estimate to the expected yield in the $6b$ region is constrained in the fit from the yields in the $5b$ Control Region. Altogether, the term $\mu_{NF} f_i(\vec{\theta}) B_i$ describes the background estimate in the $6b$ region. The second and third products are over the systematic uncertainties associated with the signal and background, constrained by Gaussian distributions with means of 0 and standard deviations of 1. The final product over all bins of continuous variables P_T represents the statistical uncertainty in MC by constraining the NPs $\vec{\gamma}$. The $\vec{\nu}$ are terms representing the effective number of MC events having taken the statistical uncertainty into account.

The very-low DNN score region, referred to as the Excluded region in Table 4.8, is excluded from the fit due to the poor background modelling. The binning schemes of the DNN score used in the fit for each analysis interpretation are chosen to maximise signal sensitivity while maintaining the validity of the systematic uncertainty modelling in the background estimate. Studies explore equal bin width as well as unequal bin width that yielded monotonically increasing Z significance in each bin from the Low-score to High-score region. The final fit, for the resDNN, is performed over bins of equal width 0.05 between $0.05 < \text{score} < 1.0$. For the nonresDNN, the fit is performed over bins of equal width 0.05 between $0.1 < \text{score} < 0.95$, and bins of equal width 0.025 between $0.95 < \text{score} < 1.0$. For the heavyresDNN, the fit is performed over bins of equal width 0.05 between $0.1 < \text{score} < 1.0$.

4.9 Results

4.9.1 HHH Production Cross-section Upper Limits

The upper limits on the production cross-section at the 95% CL are extracted by performing profile likelihood fits on the DNN scores for the various signal hypotheses at a reference production cross-section of $\sigma_{ref} = 50$ fb. The upper limits are calculated by multiplying the fit POI, the signal strength μ_{sig} , with σ_{ref} . For the non-resonant interpretation, the nonresDNN score distributions of the SM HHH signal and a non-resonant TRSM signal with $(m_X, m_S) = (400, 200)$ GeV are shown in Figure 4.21. For the resonant interpretation, the resDNN score distribution of a resonant TRSM signal with $(m_X, m_S) = (500, 350)$ GeV is shown in Figure 4.22. For the heavy-resonant interpretation, the heavyresDNN score distributions of heavy resonance signals with $(m_X, m_S) = (900, 325)$ GeV with narrow and wide decay width are shown in Figure 4.23. The 30 most impactful nuisance parameters (NPs) in the fits are ranked by their post-fit impact in Figures 4.24, 4.25 and 4.26 for each of the fits shown. The post-fit impacts are often smaller than the pre-fit impacts, as the NPs are constrained in the fit with the additional information provided in the observed data. In the same rankings, the NP pulls $(\hat{\theta} - \theta_0)/\Delta\theta$ are shown, which are the differences between the best-fit values of the NPs $\hat{\theta}$ and the nominal pre-fit values θ_0 divided by their pre-fit uncertainties

$\Delta\theta$. This non-closure between the nominal and best-fit values indicates the amount the NP needs to be varied to match the observed data in the fit. In general, a pull of larger than 1σ is considered large, especially for statistically limited cases where the statistical fluctuations can compensate for large differences. Overall for each interpretation, the systematic uncertainties associated with the background estimate are the most dominant NPs. Their variations across the DNN scores are documented in Appendix E.

For all three interpretations, some non-closure is observed between the data and the pre-fit background distributions. In the non-resonant and resonant interpretations, an excess in data is observed in the region of $0.55 < \text{score} < 0.7$ and $0.6 < \text{score} < 0.85$ of the nonresDNN respectively. In the case of the heavy-resonant interpretation, a deficit is observed in data compared to the pre-fit background estimate in $0.85 < \text{score} < 0.1$ of the heavyresDNN. However, after fitting with the background-only hypothesis, good agreement is observed between data and the post-fit background distributions for all three interpretations. This is largely due to the non-closure in the systematic uncertainties of the background estimate. For example, for the resonant interpretation, the systematic uncertainty on the background from the $H_{T_{6jets}}$ variable is pulled by more than $+1\sigma$ from its nominal values to achieve this good post-fit agreement. This is due to the largest variation in this systematic uncertainty coinciding with the score region where an excess in data compared to the pre-fit background is observed, as seen in Figure E.2(c).

The expected and observed upper limits of HHH production via the non-resonant and resonant TRSM production interpolated in the (m_X, m_S) space are shown in Figure 4.27. The upper limits for the generic heavy resonance signal with a narrow and a wide decay width are interpolated in the (m_X, m_S) space separately, shown in Figure 4.28. The expected (observed) upper limits range between 46 – 350 fb (48 – 310 fb) for the TRSM and DM-CPV models, and between 4.7 – 69 fb (5.7 – 38 fb) for production via generic heavy resonances with a narrow decay width, and between 5.2 – 53 fb (6.3 – 39 fb) for ones with a wide decay width. The upper limits of HHH production via heavy resonances are much lower than other processes, largely due to the better signal and background discrimination of the heavyresDNN. This is attributed to the larger mass of the heavier

resonance particle X in the simulated signals resulting in more boosted Higgs boson candidates. Jets produced from their decay are more collimated and therefore have smaller ambiguity in their pairing assignment, ultimately resulting in more discriminating reconstructed kinematic variables for the DNN to exploit. Overall, good agreement between the data and the background is observed. For TRSM signals in the resonant interpretation, the signal strength is exactly zero as a result of the statistical testing. For TRSM signals in the non-resonant interpretation, the largest deviation corresponding to a local significance of 0.19σ is observed for $(m_X, m_S) = (550, 200)$ GeV. For the heavy resonance signals, the largest deviation corresponding to a local significance of 0.51σ is observed for $(m_X, m_S) = (1500, 275)$ GeV.

The expected (observed) upper limits on the SM HHH signal strength are 750 (760) times the predicted SM signal strength, corresponding to an expected (observed) upper limit on the cross-section of 59 fb (59 fb) at the 95% confidence level.

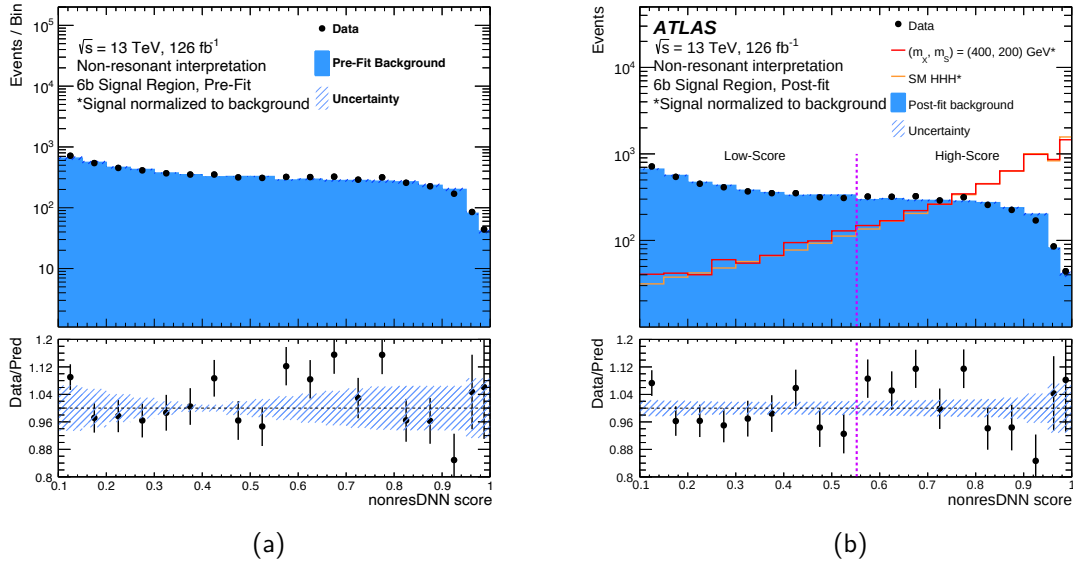


Figure 4.21: Pre-fit (a) and post-fit (b) distributions of scores for nonresDNN of $6b$ data and the background prediction. In the post-fit distributions, the boundary between the Low- and High-Score regions is shown by a dashed vertical line. SM HHH and non-resonant TRSM signals with $(m_X, m_S) = (400, 200)$ GeV are overlaid in the post-fit distributions. The lower panel in each plot shows the ratio of the data and the pre-fit or post-fit background (Pred). The uncertainty on the background prediction is shown as a hatched band centred on unity. The Low- and High-score regions are defined such that the former has less than 10% signal contamination and the latter contains the most signal-like events. The post-fit figure is taken from [1].

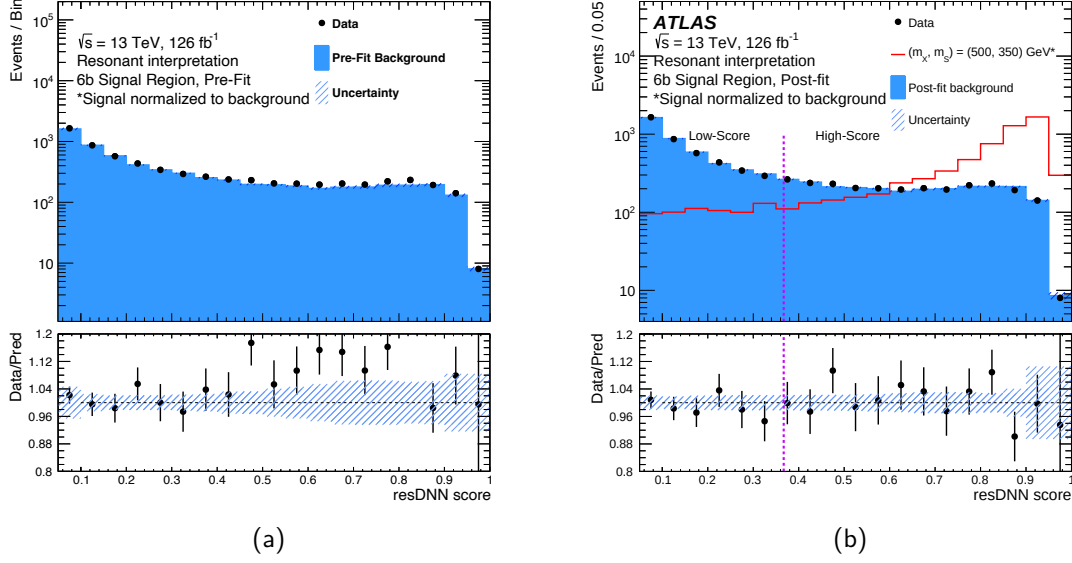


Figure 4.22: Pre-fit (a) and post-fit (b) distributions of scores for resDNN of $6b$ data and the background prediction. In the post-fit distributions, the boundary between the Low- and High-Score regions is shown by a dashed vertical line. Resonant TRSM signal with $(m_X, m_S) = (500, 350)$ GeV (normalised to the background prediction) is overlaid in the post-fit distributions in 4.22(b). The lower panel in each plot shows the ratio of the data and the pre-fit or post-fit background (Pred). The uncertainty on the background prediction is shown as a hatched band centred on unity. The Low- and High-score regions are defined such that the former has less than 10% signal contamination and the latter contains the most signal-like events. The post-fit figure is taken from [1].

4.9.2 Upper Limits on κ_3 and κ_4

Constraints on the coupling modifiers κ_3 and κ_4 are placed simultaneously while assuming SM values for all other couplings in the non-resonant interpretation. The upper limits extracted by performing a likelihood scan over the profile likelihood ratio $-\ln \Delta L$, where the two coupling modifiers are simultaneously varied using the statistical fitting setup outlined in Section 4.8.4 are shown in Figure 4.29. The unitarity bounds are placed to preserve perturbative unitarity in the $HH \rightarrow HH$ scattering at tree-level, discussed in [99]. At the 95% confidence level, no values of the coupling modifiers are excluded within the unitarity bounds. Assuming the SM value of the trilinear Higgs self-coupling, that is if $\kappa_3 = 1$, κ_4 is constrained to be between $-230 - 240$ at 95% confidence level. While assuming the SM value of the quartic Higgs self-coupling, that is if $\kappa_4 = 1$, κ_3 is constrained to be between $-11 - 17$ at the 95% confidence level.

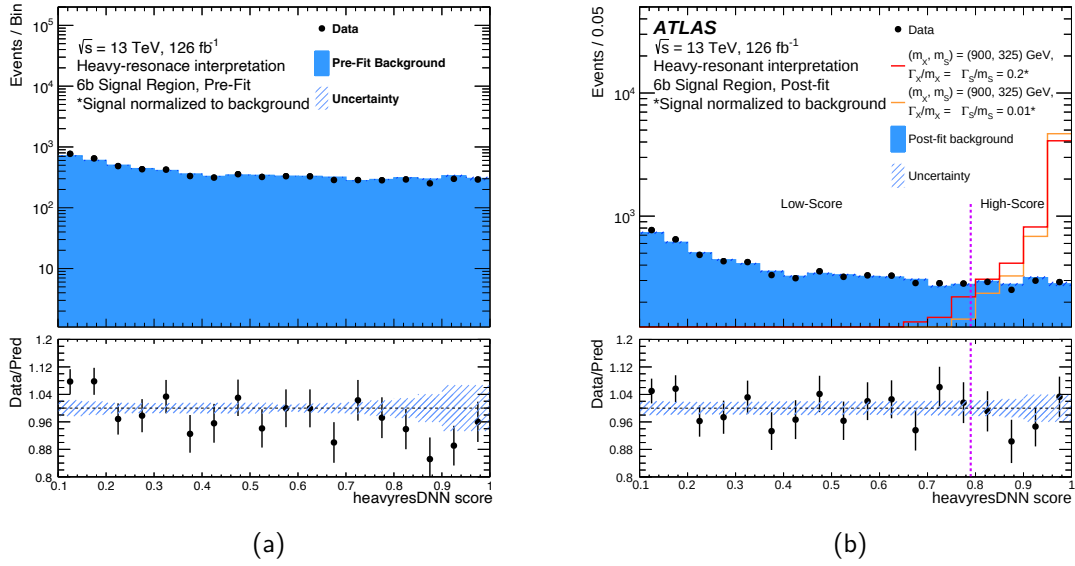


Figure 4.23: Pre-fit (a) and post-fit (b) distributions of scores for heavyresDNN of $6b$ data and the background prediction. In the post-fit distributions, the boundary between the Low- and High-Score regions is shown by a dashed vertical line. Narrow- and a wide-width heavy resonances with $(m_X, m_S) = (900, 325)$ GeV (normalised to the background prediction) are overlaid in the post-fit distributions in (b). The lower panel in each plot shows the ratio of the data and the pre-fit or post-fit background (Pred). The uncertainty on the background prediction is shown as a hatched band centred on unity. The Low- and High-score regions are defined such that the former has less than 10% signal contamination and the latter contains the most signal-like events. The post-fit figure is taken from [1].

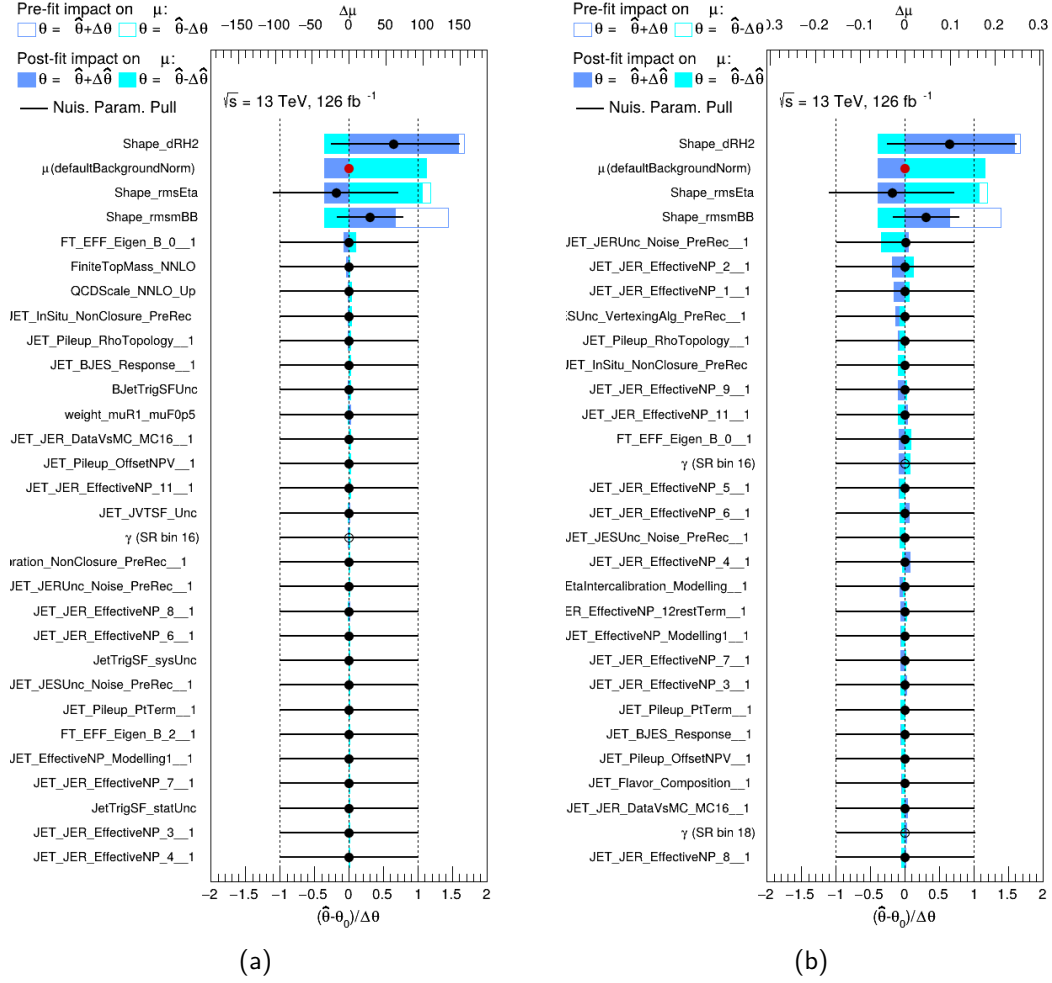


Figure 4.24: The pre- and post-fit impacts of the 30 most impactful NPs in the non-resonant interpretation fit for the SM HHH signal (a), and a non-resonant TRSM signal with $(m_X, m_S) = (400, 200)$ GeV (b), ranked in descending order of post-fit impact. The pre-fit impacts are defined as the changes in the POI when a given NP is fixed to $\theta \pm \sigma_\theta$, and the post-fit impacts are the changes in POI when it is fixed to $\hat{\theta} \pm \hat{\sigma}_\theta$. The upper x-axis $\Delta\mu$ indicates the change in the POI, and the lower x-axis $(\hat{\theta} - \theta_0)/\Delta\theta$ indicates the size of the pull on the NPs. The NPs representing the MC statistical uncertainty in a given DNN score bin are γ . The NPs representing the systematic uncertainties associated with the shape of the background estimate are labelled with the prefix Shape_. Ones representing the systematic uncertainties in jet reconstruction are labelled with the prefix JET_, and ones associated with b -tagging have the prefix FT_. NPs associated with systematic uncertainties from theory modelling have prefixes PDF_, QCD_Scale_ and FiniteTopMass_. JetTrigSF_sysUnc and BJetTrigSFUnc represent the NPs associated with uncertainties on the jet and b -jet trigger scale factors respectively. The uncertainty on the freely floating background normalisation factor $\mu(\text{defaultBackgroundNorm})$ reflects in the statistical uncertainty in the $5b$ data used to constrain it.

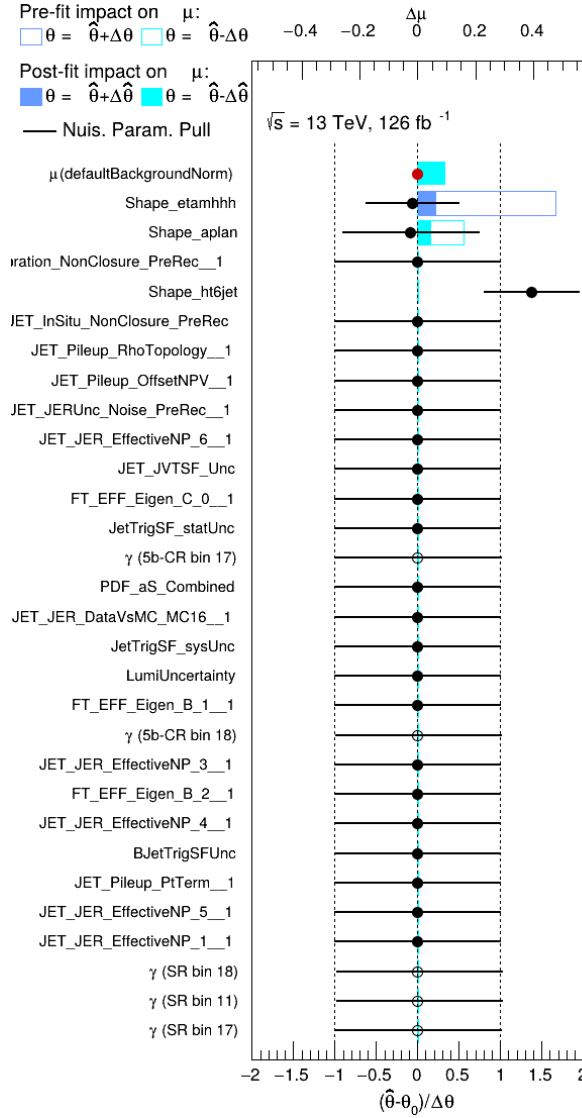


Figure 4.25: The pre- and post-fit impacts of the 30 most impactful NPs in the resonant interpretation fit for the TRSM signal with $(m_X, m_S) = (900, 325)$ GeV, ranked in descending order of post-fit impact. The pre-fit impacts are defined as the changes in the POI when a given NP is fixed to $\theta \pm \sigma_\theta$, and the post-fit impacts are the changes in POI when it is fixed to $\hat{\theta} \pm \hat{\sigma}_\theta$. The upper x-axis $\Delta\mu$ indicates the change in the POI, and the lower x-axis $(\hat{\theta} - \theta_0)/\Delta\theta$ indicates the size of the pull on the NPs. The NPs representing the MC statistical uncertainty in a given DNN score bin are γ . The NPs representing the systematic uncertainties associated with the shape of the background estimate are labelled with the prefix Shape_. Ones representing the systematic uncertainties in jet reconstruction are labelled with the prefix JET_, and ones associated with b -tagging have the prefix FT_. NPs associated with systematic uncertainties from theory modelling have prefixes PDF_, QCD_Scale_ and FiniteTopMass_. JetTrigSF_sysUnc and BJetTrigSFUnc represent the NPs associated with uncertainties on the jet and b -jet trigger scale factors respectively. The uncertainty on the freely floating background normalisation factor $\mu(\text{defaultBackgroundNorm})$ reflects in the statistical uncertainty in the $5b$ data used to constrain it.

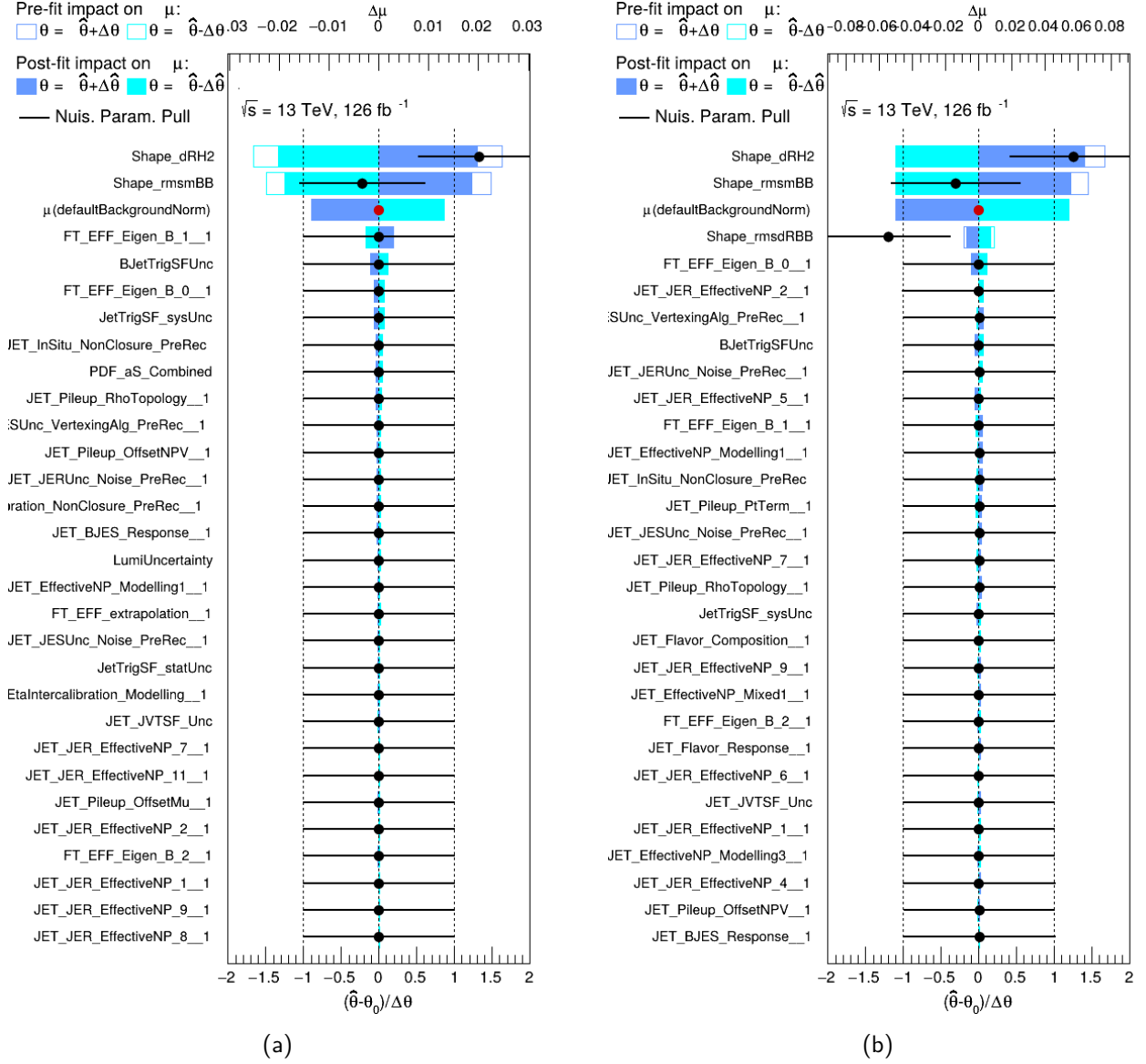


Figure 4.26: The pre- and post-fit impacts of the 30 most impactful NPs in the heavy-resonant interpretation fit for signals with $(m_X, m_S) = (900, 325)$ GeV with narrow (a) and wide (b) decay widths, ranked in descending order of post-fit impact. The pre-fit impacts are defined as the changes in the POI when a given NP is fixed to $\theta \pm \sigma_\theta$, and the post-fit impacts are the changes in POI when it is fixed to $\hat{\theta} \pm \hat{\sigma}_\theta$. The upper x-axis $\Delta\mu$ indicates the change in the POI, and the lower x-axis $(\hat{\theta} - \theta_0)/\Delta\theta$ indicates the size of the pull on the NPs. The NPs representing the MC statistical uncertainty in a given DNN score bin are γ . The NPs representing the systematic uncertainties associated with the shape of the background estimate are labelled with the prefix Shape_. Ones representing the systematic uncertainties in jet reconstruction are labelled with the prefix JET_, and ones associated with b -tagging have the prefix FT_. NPs associated with systematic uncertainties from theory modelling have prefixes PDF_, QCD_Scale_ and FiniteTopMass_. JetTrigSF_sysUnc and BJetTrigSFUnc represent the NPs associated with uncertainties on the jet and b -jet trigger scale factors respectively. The uncertainty on the freely floating background normalisation factor $\mu(\text{defaultBackgroundNorm})$ reflects in the statistical uncertainty in the $5b$ data used to constrain it.

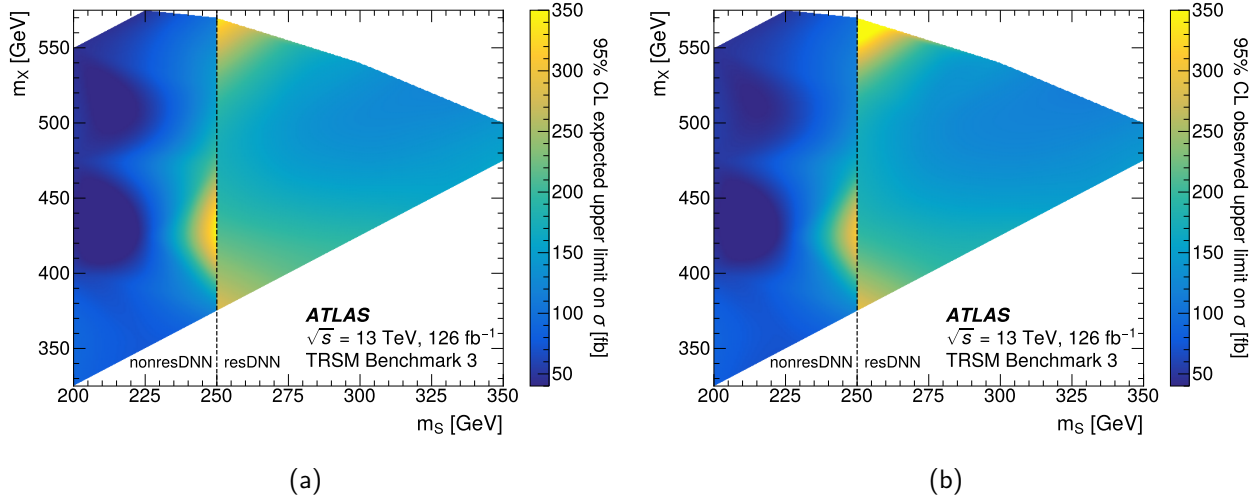


Figure 4.27: The expected (a) and observed (b) HHH cross section upper limits at the 95% CL for the phase space within the perturbative unitarity bounds of the TRSM. While the points were generated under Benchmark point 3 of the TRSM, they can also be interpreted in the DM-CPV model. A cubic B  zier polynomial is used to interpolate across the plane. The figures are taken from [1].

4.10 Conclusions

The search for HHH production in the $6b$ final state is conducted in this analysis using 126 fb^{-1} of the ATLAS Run-2 dataset from proton-proton collisions at the LHC. The results are interpreted in three different production topologies, resonant, non-resonant and heavy-resonant. Upper limits are placed on the production cross-sections of the SM HHH and various BSM models, namely via the TRSM, DM-CPV models and generic heavy scalars. The non-resonant interpretation also searches for SM-like HHH production with modified values of the trilinear and quartic Higgs self-coupling constants, characterised by the corresponding modifiers κ_3 and κ_4 . No significant evidence of HHH signals via the SM or BSM HHH productions has been observed. Simultaneous limits are placed on the trilinear and quartic Higgs self-coupling modifiers, making them the first ever experimental constraints on κ_4 in relation to κ_3 . At the 95% confidence level, assuming $\kappa_4 = 1$, κ_3 is constrained between $-11 - 17$, whereas assuming $\kappa_3 = 1$, κ_4 is constrained between $-230 - 240$, for both the expected and observed limits. For reference, the combined searches for HH production, dependent on the trilinear Higgs self-coupling, are able to constrained κ_3 to be between $-1.2 - 7.2$ at 95% CL [100]–[102]. The more recent combination of single- and di-Higgs production measurements at ATLAS further constrains κ_3 to be between $-0.4 - 6.3$ [14].

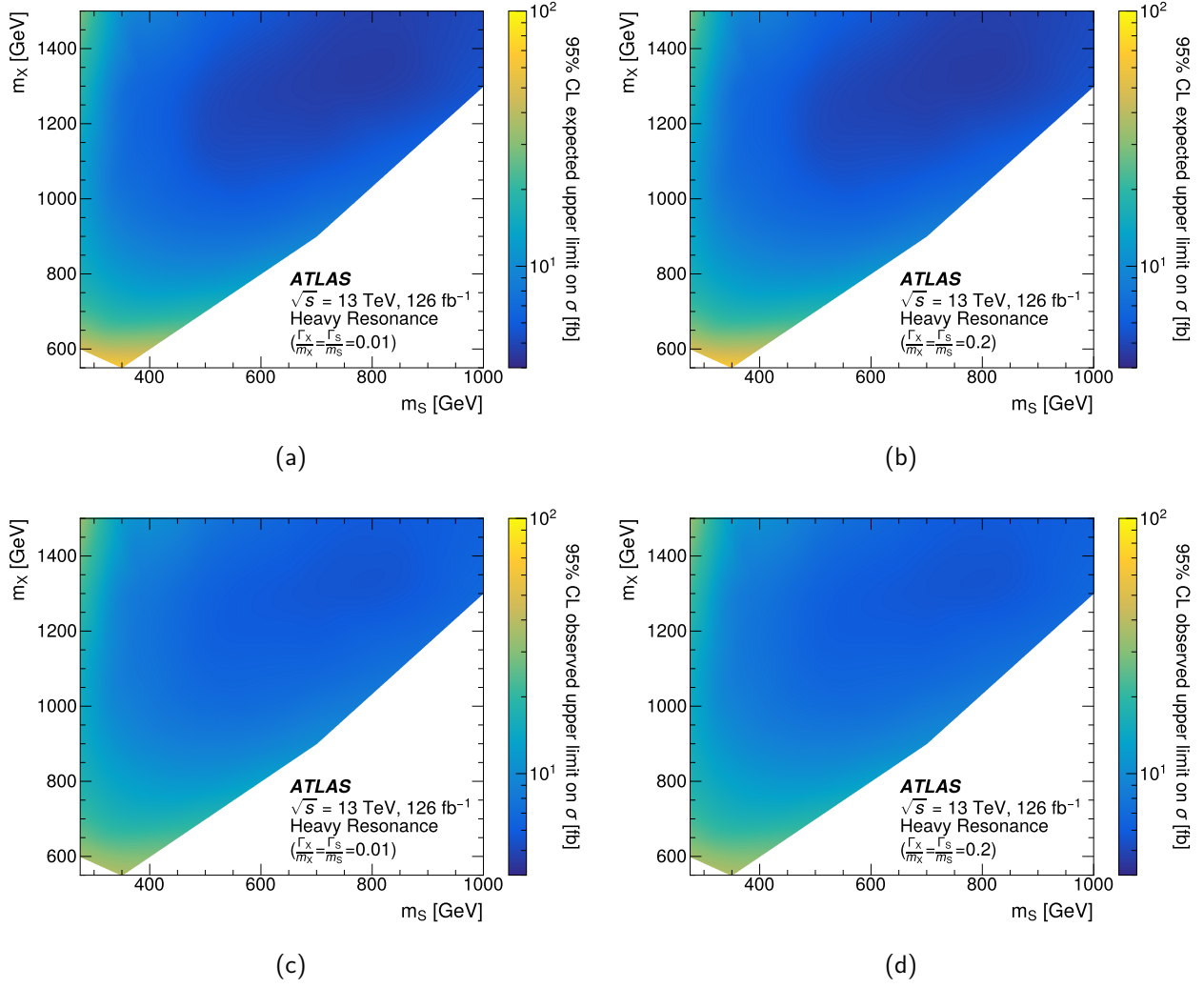


Figure 4.28: The expected (a), (b) and observed (c), 4.28(d) HHH cross section upper limits at the 95% CL for the heavy resonance signals with narrow or wide decay widths. A cubic B  zier polynomial is used to interpolate across the plane.

For the BSM productions, upper limits are placed on the various allowed production cross-sections of the targeted models. An expected (observed) upper limit of the SM HHH signal strength is placed at 750 (760) times the SM prediction, corresponding to an upper limit of an expected (observed) upper limit on the cross-section of 59 fb (59 fb) at the 95% confidence level. The expected (observed) upper limits between 46 – 350 fb (48 – 310 fb) are placed for production cross-sections via the TRSM and DM-CPV models at the 95% confidence level. For productions via generic heavy scalars, the expected (observed) upper limits at the 95% confidence level on the production cross-sections range between 4.7 – 69 fb (5.7 – 38 fb) for resonances with a narrow decay width, and between 5.2 – 53 fb (6.3 – 39 fb) for the wide decay width.

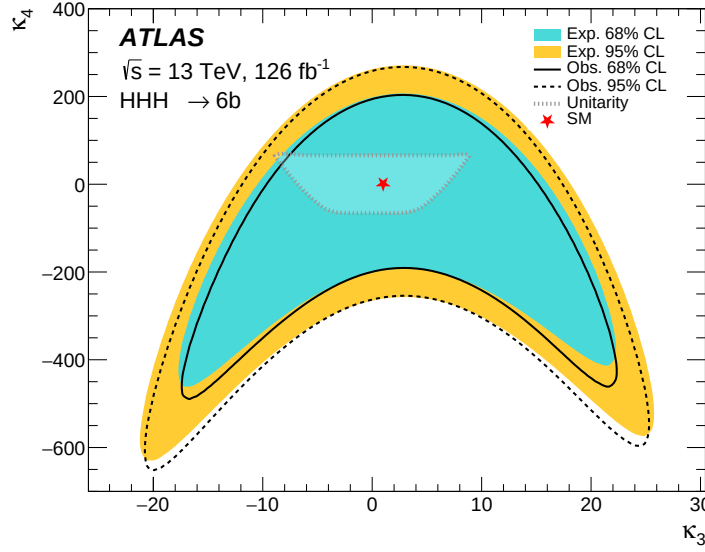


Figure 4.29: Expected (filled regions) and observed (black solid and dashed lines) 95% and 68% CL constraints on the Higgs trilinear and quartic self-coupling modifiers κ_3 and κ_4 . Unitarity limits, calculated in [99], are overlaid in the region bounded by the gray dashed line. The red star is the SM values of the coupling modifiers, $\kappa_3 = \kappa_4 = 1$. This figure is taken from [1].

4.11 Outlooks

The analysis presents the first ever search for HHH production at the LHC. It sets the foundation for multi-Higgs analyses in the future, with the goals of providing more powerful constraints on the Higgs self-coupling, an important free parameter in the SM as well as probing BSM models in the extended Higgs sectors. Beyond this search using Run-2 data, the ever growing Run-3 dataset will provide valuable statistics for these endeavours. Numerous efforts are being initiated to further probe HHH production in not only the $6b$ but also various other final states. For example, the higher signal to background ratio provided by one of the Higgs bosons decay to a pair of τ -leptons, as well as the recent advancement in its identification offer a trade-off with the reduced branching that might be feasible for a future analysis using the Run-3 dataset.

Looking further ahead, the prospect of searching for the non-resonant HHH is projected for the High-luminosity LHC with a center-of-mass energy of $\sqrt{s} = 14 \text{ TeV}$ and an integrated luminosity ranging between $1000 - 3000 \text{ fb}^{-1}$ in [103]. In the scenario including conservative systematic uncertainties, the expected sensitivity to the SM HHH signal strength reaches 109 times the SM

prediction at 95% confidence level. The expected constraints on the quartic Higgs self-coupling modifier, assuming SM value for the trilinear modifier, are projected to range between -84, 96. The projected simultaneous constraints placed on κ_3 and κ_4 at the 95% confidence level for an integrated luminosity 3000 fb^{-1} using various extrapolation scenarios are shown in Figure 4.30. These scenarios not only account for the large increase in dataset in the future, but crucially also the improvements in flavour-tagging algorithms and reduced uncertainties in the background estimation. These projections provide an exciting outlook on our ability to probe the Higgs self-coupling, therefore gaining more knowledge about the shape of the Higgs potential, the nature of the electroweak phase transition and the metastability of the Universe.

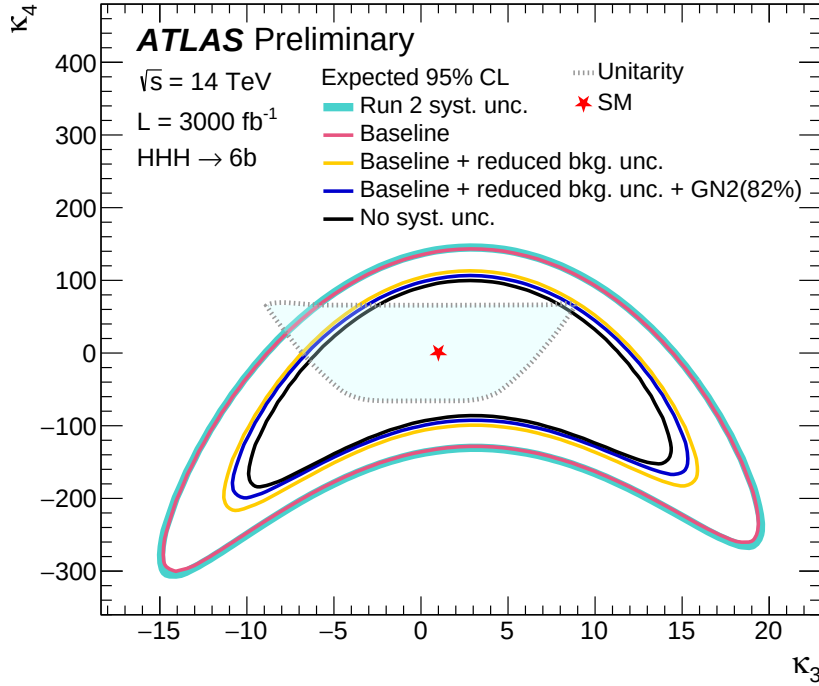


Figure 4.30: The expected constraints on the trilinear and quartic Higgs self-coupling modifiers κ_3 and κ_4 projected for the HL-LHC. The constraints are shown for an integrated luminosity of 3000 fb^{-1} , comparing different extrapolation scenarios including reduced background shape uncertainties and improved b -tagging efficiency with unchanged background mis-tagging efficiencies. The projected constraints are compared with the ones placed using the ATLAS Run-2 dataset. The figure is taken from [103].

5 | Alternative Strategies for $HHH \rightarrow 6b$

A good tomme will outdo a good morbier, but a
decent morbier will outdo a cheap tomme.
— Bethany Spear

During the course of the $HHH \rightarrow 6b$ analysis, extensive exploration of various analysis strategies was carried out, given it was the first ever search of such topology at the LHC. Although most of the methods ultimately were not adopted in the final version of the analysis - such is the nature of research, they are nevertheless interesting studies in their own rights and may provide inspirations for multi-Higgs analyses in the future.

5.1 Event Categorisation in the m_H Volume

For a signature that contains three SM Higgs bosons, an efficient way of separating signal from background events is to reconstruct their individual masses. As only signal events would contain three genuine Higgs bosons, the reconstructed m_H values are expected to be close to the mass of the SM Higgs boson at 125 GeV. Whereas in the background-enriched events, their distributions are expected to have no distinctive peaks around 125 GeV. Alternative definitions of the analysis regions in the (m_{H1}, m_{H2}, m_{H3}) volume are therefore explored for better signal and background separation, detailed in Section 5.1.1. However, the reconstruction of m_H values requires explicit jet pairing, with 15 unique ways of assigning 6 jets to 3 Higgs boson candidates. In order to maximise the pairing efficiency for the signal events, a machine learning approach using a Deep Neural Network (DNN) is explored to predict the correct pairing combination for signal events, outlined in Section 5.1.2. To further reduce the number of background events retained in the Signal Region in the (m_{H1}, m_{H2}, m_{H3}) volume, a DNN classifier is trained and applied to only events in this region, outlined in Section 5.1.3. The advantages and disadvantages are discussed in the following sections, which justify the final chosen strategy for the $HHH \rightarrow 6b$ analysis.

5.1.1 Analysis Regions

The mass of the Higgs boson candidates can be reconstructed using the baseline jet pairing algorithm that minimises the quantity $|m_{H1} - 120 \text{ GeV}| + |m_{H2} - 115 \text{ GeV}| + |m_{H3} - 110 \text{ GeV}|$. The masses of the three Higgs boson candidates form a 3-dimensional volume in (m_{H1}, m_{H2}, m_{H3}) . The point $(m_{H1}, m_{H2}, m_{H3}) = (120, 115, 110) \text{ GeV}$ will be referred to as the pairing centre. Due to the relatively symmetric distributions of events in this volume, a spherical Signal Region with a 30 GeV radius centred around the pairing centre is defined to retain most of the signal events. Surrounding it, spherical shells of radii 20 GeV and 15 GeV enriched by background events are defined as the Control and Validation Regions, which can be used to model the background events inside the Signal Region. The dimensions of these regions are chosen such that a reasonable amount of the simulated $6b$ signal events fall in the Signal Region, while sufficient $5b$ data events are within the Control and Validation Regions for background modelling. The sizes of the Control and Validation Regions are restricted by the trigger thresholds imposed on the jets which limit the lower bounds of the reconstructed m_H values. The distributions of the combined $6b$ resonant TRMS signal events and the background-enriched $5b$ data events in the (m_{H1}, m_{H2}, m_{H3}) volume are shown in Figures 5.1 and 5.2, where they are projected onto the 2-dimensional planes formed by 2 out of the 3 m_H 's. The effective yields and the $\frac{S}{\sqrt{B}}$ of the resonant TRSM signal at $(m_X, m_S) = (450, 275) \text{ GeV}$ and $5b$ data events contained in each of the analysis regions are summarised in Table 5.1. The $\frac{S}{\sqrt{B}}$ represents a rough estimate of the signal sensitivity in each region. The signal sensitivity in the spherical Signal Region achieved is higher than in the inclusive region, while the sensitivity in the Control and Validation Regions are sufficiently low for the $5b$ data events to be used for background modelling. However, the sculpting in the m_H distributions of the background events is clearly evident, where most of the $5b$ data events are clustered around the pairing centre despite not containing three genuine Higgs bosons. As a result, the spherical Signal Region contains a large amount of background as well as signal events, making this analysis region definition less optimal.

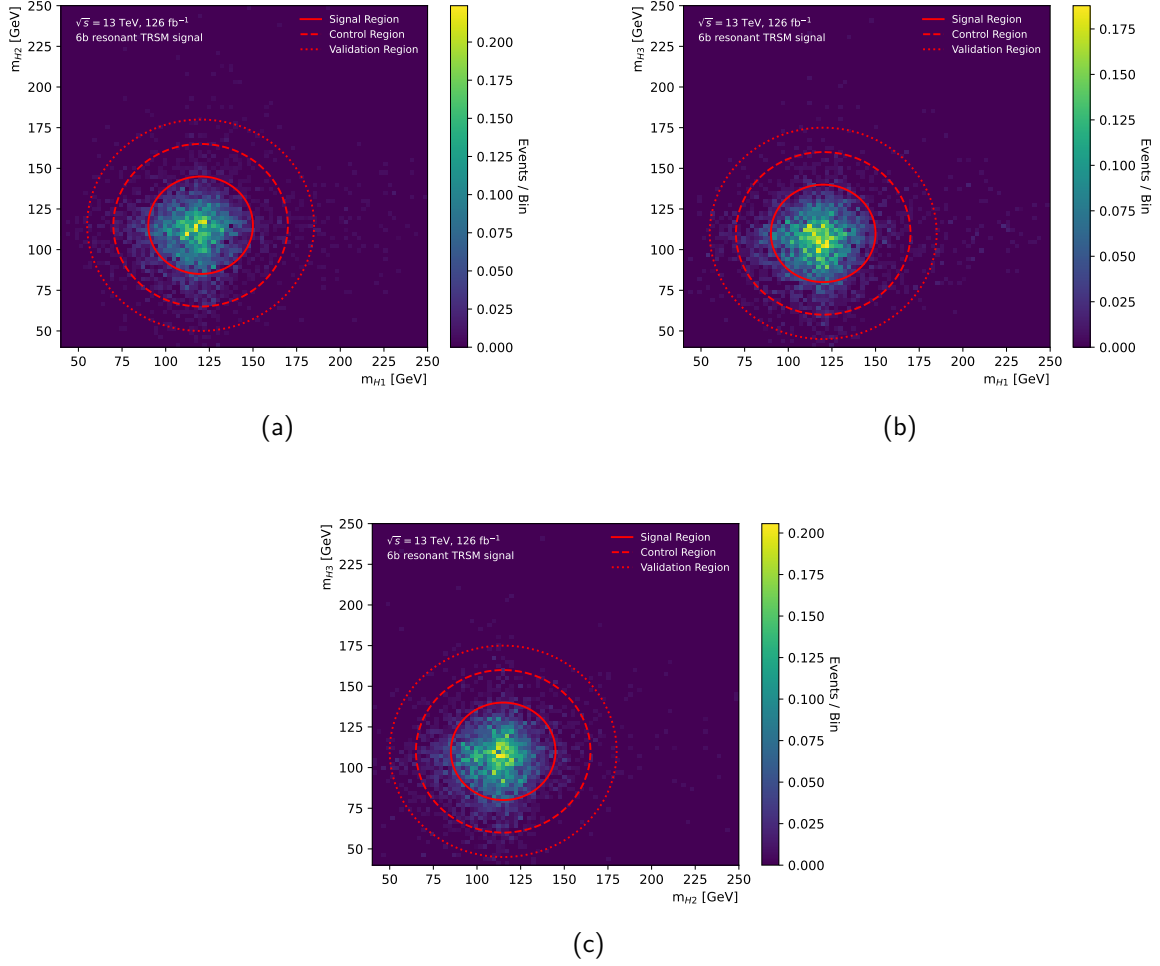
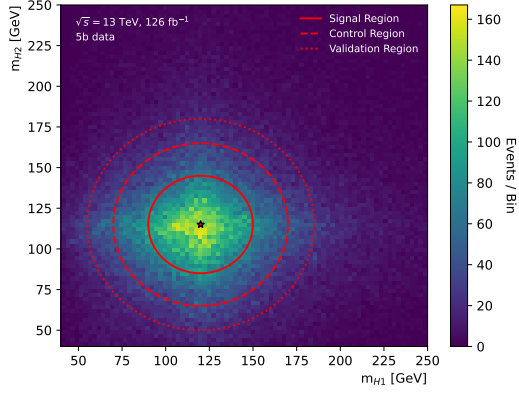
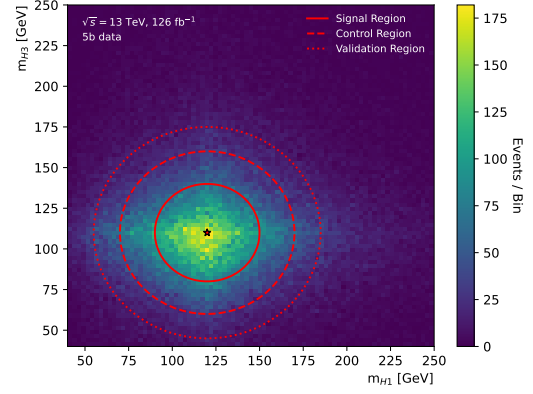


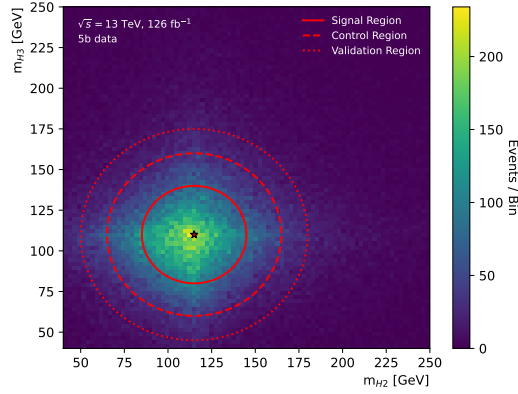
Figure 5.1: Distributions of the $6b$ events of a resonant TRSM signal at $(m_X, m_S) = (450, 275)$ GeV in the (m_{H1}, m_{H2}, m_{H3}) volume, projected onto the 2-dimensional planes formed by 2 out of the 3 Higgs boson candidate masses. The baseline jet pairing method is used, where the reconstructed Higgs candidates have masses close to the pairing centre at (120, 115, 110) GeV. A spherical Signal Region of radius 30 GeV is indicated by the red solid line, a spherical shell Control Region of radius 20 GeV and a spherical shell Validation Region of radius 15 GeV are indicated by red dashed and dotted lines respectively.



(a)



(b)



(c)

Figure 5.2: Distributions of the $5b$ data events in the (m_{H1}, m_{H2}, m_{H3}) volume, projected onto the 2-dimensional planes formed by 2 out of the 3 Higgs boson candidate masses. The baseline jet pairing method is used to reconstruct the m_H 's, where the pairing centre is marked by the red star. Centred around the pairing centre, a spherical Signal Region of radius 30 GeV is indicated by the red solid line, a spherical shell Control Region of radius 20 GeV and a spherical shell Validation Region of radius 15 GeV are indicated by red dashed and dotted lines respectively.

	Inclusive	Signal Region	Control Region	Validation Region
$6b$ resonant TRSM signal	45.1	21.8	10.9	2.1
$5b$ data	141428	23099	40026	25707
$4b$ data	1937723	317852	536033	343942
$\frac{S}{\sqrt{B}}$	0.44	0.53	0.20	0.048

Table 5.1: Table summarising the effective yields and the $\frac{S}{\sqrt{B}}$ of the $6b$ events of a resonant TRSM signal at $(m_X, m_S) = (450, 275)$ GeV, $5b$ and $4b$ data events inside the Signal, Control and Validation Regions defined in the (m_{H1}, m_{H2}, m_{H3}) volume, where the m_H 's are reconstructed using the baseline pairing method. The "Inclusive" column shows the total number of events in each sample. The $\frac{S}{\sqrt{B}}$ represents a rough estimate of the signal sensitivity over the expected $6b$ background, which is obtained by scaling the $5b$ data yields by the ratio between $5b$ to $4b$ data.

5.1.2 Alternative Jet Pairing Methods

An alternative jet pairing method is explored to reduce the sculpting effect in the m_H distributions but yields a lower pairing efficiency. The method uses a DNN, trained on the combined $6b$ resonant TRSM signal events to predict the correct pairing configuration of the 6 b -tagged jets with leading p_T in each event. The DNN performs binary classification between events reconstructed using the correct pairing and the incorrect pairing combinations. During training, the correct pairing configurations are assigned with a target label of 1, and the incorrect ones assigned with a target label of 0. A pairing score between 0 and 1 is provided by the DNN for each configuration. When applied to data, all 15 possible configurations of pairing for each event are provided to the model, and the one with the highest pairing score is selected. The distributions of the network outputs evaluated on the combined $6b$ resonant TRSM signal events in the training and validation datasets are shown in Figure 5.3. The discrimination between correct and incorrect pairing configurations is reasonably good, but the false positive rate in the high-score region is relatively large.

The average pairing efficiency across all the signals in the resonant TRSM samples is 41%, compared to the average efficiency of 50% achieved by the baseline pairing method. The lower pairing efficiency is also evident in the distribution of the $6b$ events of one of the resonant TRSM signal samples at $(m_X, m_S) = (450, 270)$ GeV, shown in Figure 5.4, where the m_H values of signal events are less clustered around the SM Higgs boson mass. However, the sculpting of the background events is significantly less severe, such that larger fractions of $5b$ data events lie in the Control and Validation

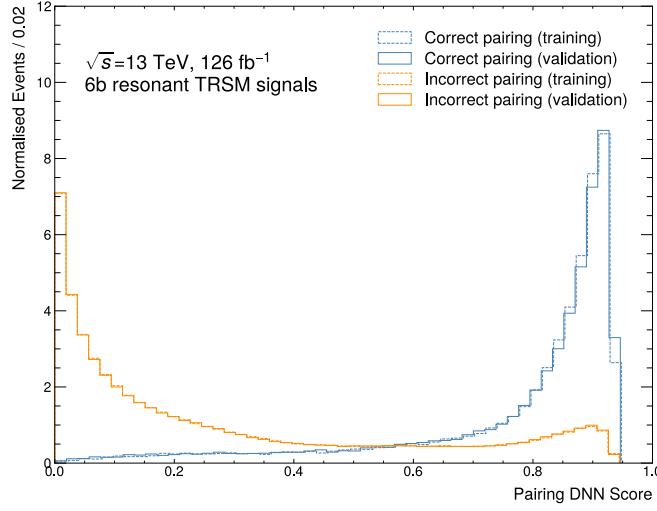


Figure 5.3: Output score distributions of the jet pairing DNN evaluated on the training and validation datasets of the combined $6b$ resonant TRSM signals.

Regions in Figure 5.5. The effective yields and the $\frac{S}{\sqrt{B}}$ of the resonant TRSM signal at $(m_X, m_S) = (450, 275)$ GeV and $5b$ data events contained in each of the analysis regions are summarised in Table 5.2. Whilst achieving a similar $\frac{S}{\sqrt{B}}$ in the Signal Region as the baseline jet pairing method, the DNN-based jet pairing method is suboptimal for several reasons. The signal leakage into the Control and Validation Regions is not negligible, resulting in a signal sensitivity that is not significantly lower than the Signal Region, which undermines the validity of the background estimate derived from these regions. The available statistics of $5b$ data events in these regions are also significantly lower for the derivation and validation of this background estimate. As a result, the $HHH \rightarrow 6b$ analysis chooses a mass-based jet pairing method and a non-mass-based variable, the DNN score, as the main discriminant for the statistical analysis, which effectively avoids issues arising from the sculpting of the background distributions from the jet pairing method. However, this study shows that a ML-based jet pairing method can be as efficient as traditional methods, while significantly reducing the sculpting effect of background events.

5.1.3 Background Reduction Inside the Signal Region

In order to exclude more background events inside the spherical Signal Region defined in the (m_{H1}, m_{H2}, m_{H3}) volume, a DNN model (SRNN) is trained and applied specifically to events inside

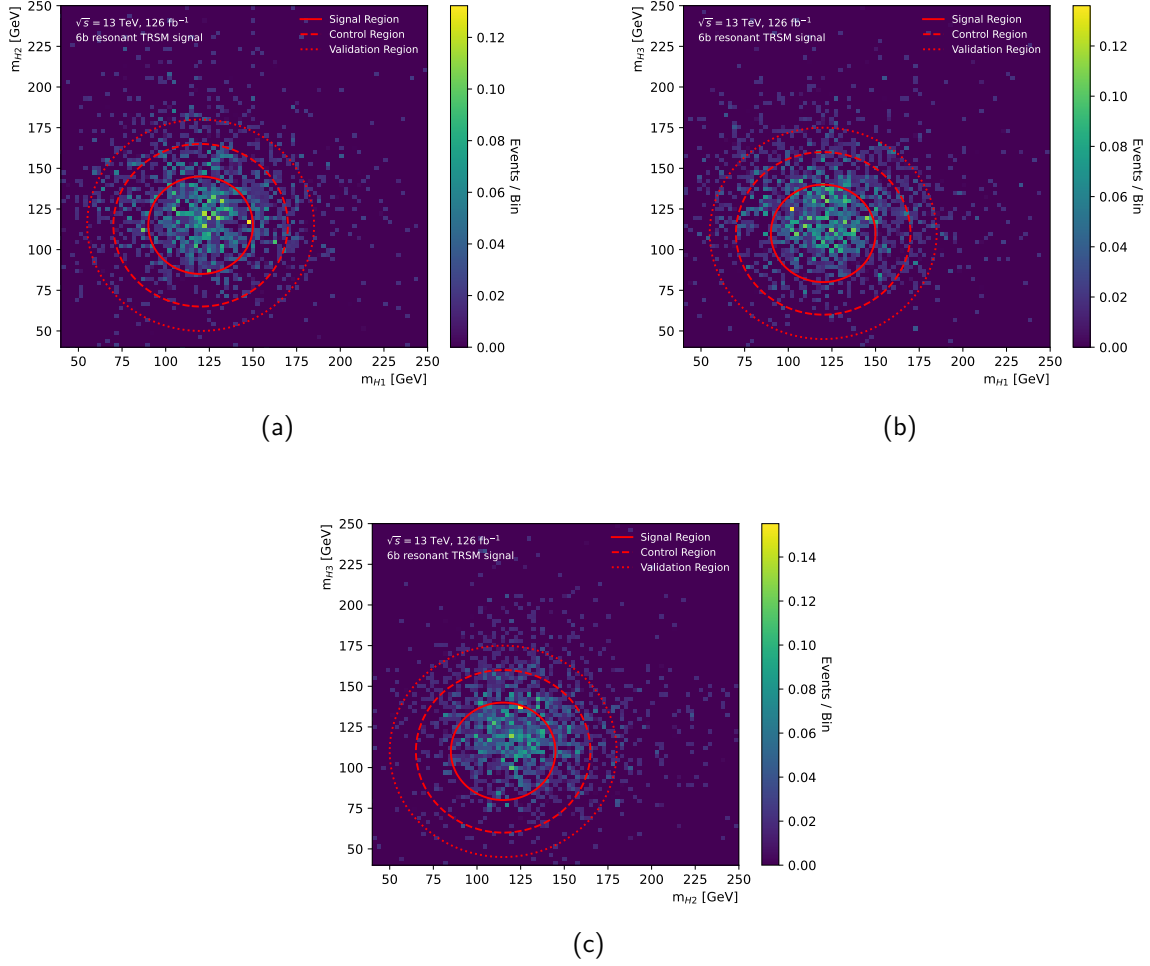


Figure 5.4: Distributions of the $6b$ events of a resonant TRSM signal at $(m_X, m_S) = (450, 270)$ GeV in the (m_{H1}, m_{H2}, m_{H3}) volume, projected onto the 2-dimensional planes formed by 2 out of the 3 Higgs boson candidate masses. The DNN-based jet pairing method is used to reconstruct the masses. Centred around $(m_{H1}, m_{H2}, m_{H3}) = (120, 115, 110)$ GeV, a spherical Signal Region of radius 30 GeV is indicated by the red solid line, a spherical shell Control Region of radius 20 GeV and a spherical shell Validation Region of radius 15 GeV are indicated by red dashed and dotted lines respectively.

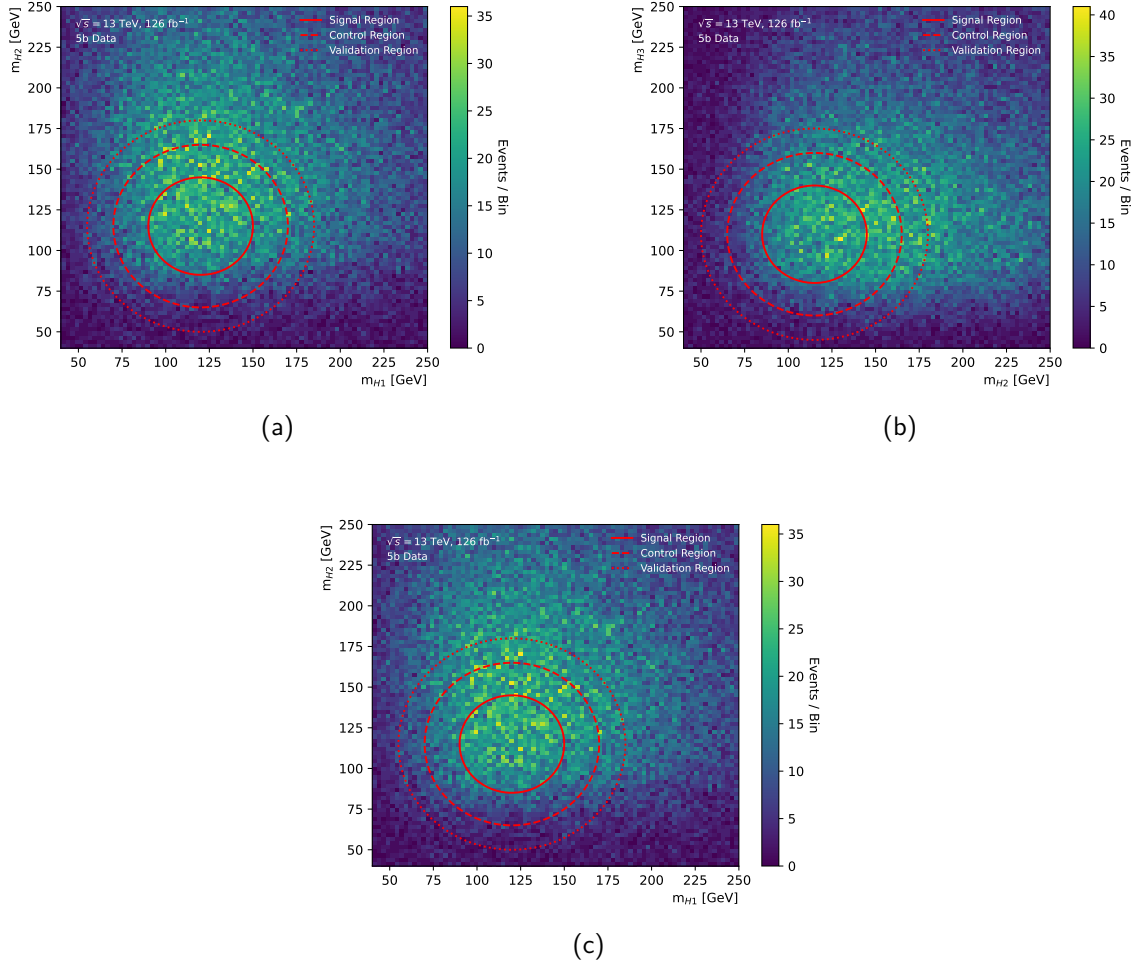


Figure 5.5: Distributions of the $5b$ data events in the (m_{H1}, m_{H2}, m_{H3}) volume, projected onto the 2-dimensional planes formed by 2 out of the 3 Higgs boson candidate masses. The DNN-based jet pairing method is used to reconstruct the masses. Centred around the pairing centre, a spherical Signal Region of radius 30 GeV is indicated by the red solid line, a spherical shell Control Region of radius 20 GeV and a spherical shell Validation Region of radius 15 GeV are indicated by red dashed and dotted lines respectively.

this region. The model uses the combined $6b$ resonant TRSM signal events as the signal class, and $5b$ data as the background, both inside the spherical Signal Region of radius 30 GeV centred at the pairing centre in the (m_{H1}, m_{H2}, m_{H3}) volume constructed using the baseline pairing method. The training procedure follows exactly as the training of the resDNN described in Section 4.4. Good signal and background separation is achieved, with an AUC of 0.889 on the validation datasets when the even and odd models are combined, shown in Figure 5.6. To maximise the number of signal events selected, a threshold on the DNN score is placed to retain 90% of the $6b$ resonant TRSM signal events, where events above it are selected for statistical fitting. As a result, roughly

	Inclusive	Signal Region	Control Region	Validation Region
$6b$ resonant TRSM signal	35.2	7.3	9.6	6.1
$5b$ data	141428	2200	6255	6588
$4b$ data	1937723	26826	75499	81291
$\frac{S}{\sqrt{B}}$	0.35	0.58	0.45	0.28

Table 5.2: Table summarising the effective yields and the $\frac{S}{\sqrt{B}}$ of the $6b$ events of a resonant TRSM signal at $(m_X, m_S) = (450, 270)$ GeV, $5b$ and $4b$ data events inside the Signal, Control and Validation Regions defined in the (m_{H1}, m_{H2}, m_{H3}) volume, where the m_H 's are reconstructed using the DNN-based jet pairing method. The "Inclusive" column shows the total number of events in each sample. The $\frac{S}{\sqrt{B}}$ represents a rough estimate of the signal sensitivity over the expected $6b$ background, which is obtained by scaling the $5b$ data yields by the ratio between $5b$ to $4b$ data.

70% of $5b$ data events are rejected. However, $5b$ data events with DNN scores above this threshold have kinematics more similar to the $6b$ resonant TRSM signal events. For example, the corrected tri-Higgs invariant mass $m_{HHH_{125}}$ in Figure 5.7 of $5b$ data events inside the Signal Region becomes more peaked in the range between 400 – 600 GeV, which is the range of the m_X values in the resonant TRSM signal samples. This decreases the signal sensitivity in a statistical fitting procedure using $m_{HHH_{125}}$ as the main discriminant. Indeed, this is seen in the higher expected upper limits on the production cross-section of a representative resonant TRSM signal sample with $m_{HHH_{125}}$ instead of the SRNN score as the main discriminant in Table 5.3. However, signal sensitivity obtained from fitting on the SRNN score is poorer than fitting on the output scores of resDNN, a model trained with events inclusively over all regions. This allows more statistics of background events in the training, albeit ones that are less signal-like and are therefore more easily separated from signal events. Fitting on the DNN score of inclusive events also allows the less signal-like events to be included to provide more information on the background estimate.

5.1.4 Conclusions

The exploration of alternative approaches to jet pairing and analysis region definition leads to the convergence to an optimal strategy for the $HHH \rightarrow 6b$ analysis. This strategy not only yields the best sensitivity to the various signal types targeted by the analysis, but is also well-motivated for physics reasons, such as the validity of the background estimation. Nevertheless, the alterna-

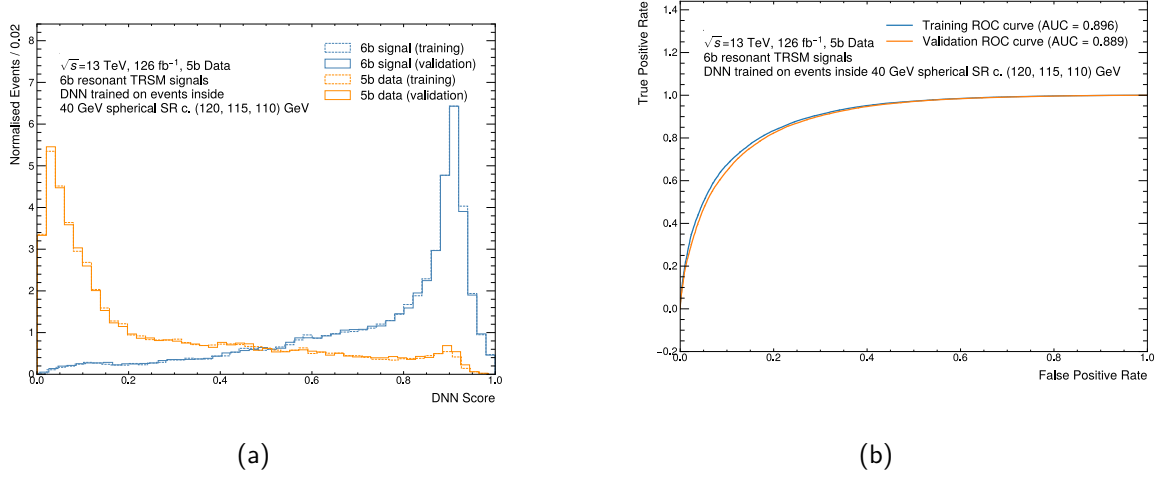


Figure 5.6: Distribution of the scores (a), and ROC curve (b) of the even and odd DNN models combined, evaluated on the combined $6b$ resonant TRSM signal events and $5b$ data events inside the spherical Signal Region of radius 30 GeV, centred at the pairing centre in the (m_{H1}, m_{H2}, m_{H3}) volume constructed using the baseline pairing method.

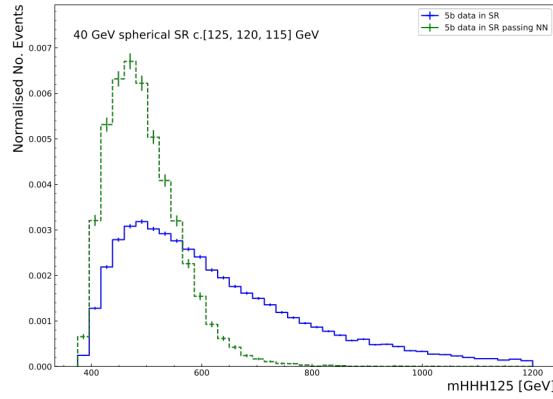


Figure 5.7: Distribution of the corrected tri-Higgs invariant mass m_{HHH125} of $5b$ data events inside the spherical Signal Region of radius 30 GeV and passing the 90% signal efficiency threshold of the SRNN.

Main discriminant	Inside Signal Region [fb]	inside Signal Region & passing SRNN threshold [fb]
m_{HHH125}	105.21	96.18
SRNN score	72.84	79.07

Table 5.3: Expected cross-section upper limits in fb for the resonant TRSM signal with $(m_X, m_S) = (450, 300)$ GeV, obtained from fitting on m_{HHH125} or the SRNN score as the main discriminant, on events inside the spherical Signal Region defined in Section 5.1, or passing the additional SRNN threshold. Asimov fits are performed, which use the derived background template from $5b$ data as the data sample, taking only the statistical uncertainty into account.

tive strategies have their own merits that could be more suitable for other signal types or physics topologies that face similar challenges. For example, the DNN-based jet pairing is promising for more complex final states with more physics objects and a larger combinatorics. It also significantly reduces the sculpting effect in the m_H distributions of background events compared to the conventional mass-based pairing method, which would be beneficial for analyses that use a mass-based variable as the main discriminant. However, it is less suitable for multi-Higgs analyses that have low p_T jets close to the trigger thresholds. This restricts the dimensions of the analysis regions, limiting the statistics available for deriving the background estimation in the Control and Validation Regions. The mass distributions of the Higgs boson candidates are also more likely to be similar to the ones reconstructed from QCD multijet events, reducing the signal and background separation in the Signal Region. To tackle the large proportion of background events included inside the spherical Signal Region due to the sculpting effect of the baseline pairing method, an additional DNN specifically trained and applied to events inside this region effectively rejects 70% background events at a threshold that retains 90% of the resonant TRSM signal events. However, background events above this threshold have more signal-like kinematics, which reduce the signal sensitivity if used as the main discriminant for the statistical fitting. The exclusion of events outside the Signal Region also reduces the available information to model the expected background distribution. For these reasons, the final strategy for $HHH \rightarrow 6b$ uses events inclusively and their DNN score as the main discriminant.

5.2 Parametrised Neural Network

Neural networks used by physics analyses often take measured parameters as input variables. However, the addition of physics parameters not only achieves better the performance without having to train separate models for each physics parameter value, but enables smooth interpolation between them. These neural networks are often referred to as Parametrised Neural Networks (pNNs), and their application in high energy physics is extensively discussed in [104]. This is particularly applicable in searches for resonant particles, where the physics parameter is the particle mass, or the truth values, used to simulate the signal samples. However, the application of such models to observed

data is not straightforward, with problems arising from the assignment of the physics parameter values to data events. Three methods of application are most common in higher energy physics:

1. Assign the same physics parameter value to all data events and evaluate the parametrised neural network for each value separately.
2. Randomly assign a parameter value used during training to each data event.
3. Assign the reconstructed parameter value to each data event.

The first application is robust for testing each physics hypothesis, but is arduous in its application, as each evaluation would require a unique background estimation. The second suffers from loss in performance, due to the non-trivial chance of assigning the wrong physics parameter value to a data event that corresponds to a genuine signal event. The third also suffers from loss in performance due to the absence of a signal that the physics parameter can be reconstructed from in most data events. These methods of application are tested using a parametrised neural network (pNN) trained on the $6b$ events of resonant TRSM HHH signal samples, using the mass of the heavier resonance m_X as the physics parameter during training. The performance is compared with a baseline DNN, namely the "Joint NN", which is trained on the combined $6b$ resonant TRSM HHH signal samples, without parametrisation.

5.2.1 Training

The training procedure closely follows the setup used to train the resDNNs in the $HHH \rightarrow 6b$ analysis, described in Section 4.4. The signal class consists of the combined $6b$ resonant TRSM signal events, and the background class consists of $5b$ data events. Events from all but three intermediate signal mass hypotheses are included in the training. The trained pNNs are applied to the excluded signal samples to evaluate their ability to interpolate between known values of the physics parameter. The input features used for the training are the same as the ones used in the resDNN training, summarised in Tables 4.6 and 4.7. The truth mass of the heavy resonance particle X is used as the physics parameter, and is included in addition to these input features during training. The model is a feedforward DNN that performs binary classification using the same binary cross-entropy loss function in Equation 4.6. Its set of hyperparameters is chosen to yield the largest

AUC, and is summarised in Table 5.4.

Parameter	pNN for resonant TRSM signals
Learning rate	0.005
Batchsize	128
Number of hidden layers	4
Number of nodes in each layer	24
Dropout rate	0.1

Table 5.4: Table summarising the optimised hyperparameters of the pairs of pNNs trained for the resonant TRSM signal events.

5.2.2 Performance Compared to the Baseline Model

Evaluated on the combined $6b$ resonant TRSM HHH signal samples, the pNN outperforms the baseline joint NN for all signal mass hypotheses included in the training, shown in Figure 5.8. For each event, the truth value of the physics parameter m_X is assigned. For the 3 mass hypotheses excluded from the training (interpolated points), the pNN yields higher AUCs than the joint NN for the two with intermediate $\frac{m_X - m_S}{m_X}$. However the AUC is lower for the sample with larger $\frac{m_X - m_S}{m_X}$. This trend is observed in for samples of signal mass hypotheses seen in training as well as for the interpolated points.

5.2.3 Parametrisation Schemes

When assigned with the truth values of the heavy resonant particle mass m_X , the pNN out-performs the baseline joint NN in almost all cases when applied to individual signal samples, seen in Figure 5.8. This is not surprising, as the training of the Joint NN is averaged over all signal samples, whereas the performance of the pNN is parametrised on the particle mass. However when the reconstructed m_X is assigned to each event instead of the truth value, the pNN performance is severely reduced, shown in Figure 5.9. This is due to the absence of a resonant particle in most data events, therefore the m_X values reconstructed from them do not correspond to any signal hypotheses.

In the case where the pNN is evaluated on events with randomly assigned parameter values, there is

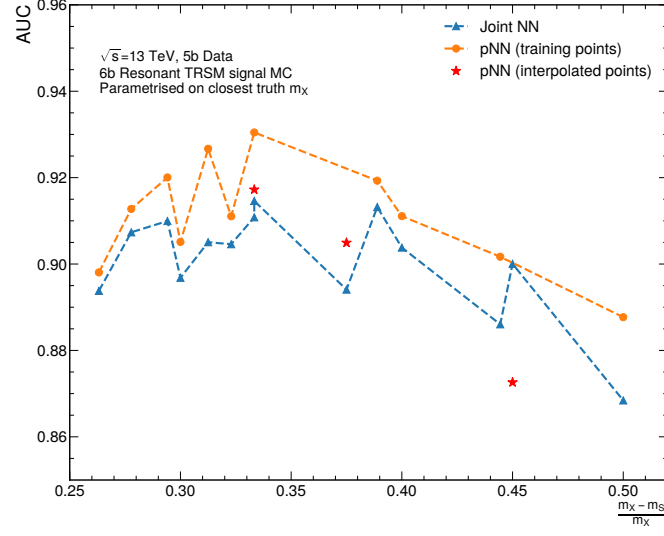


Figure 5.8: Area under the ROC curve (AUC) of the pNN and Joint NN evaluated on $6b$ resonant TRSM HHH signal samples and $5b$ data as background sample, as a function of the relative difference between the masses of the resonance particles $\frac{m_X - m_S}{m_X}$. The AUC of the pNN evaluated on three interpolated points that are excluded from the training dataset are shown in red stars.

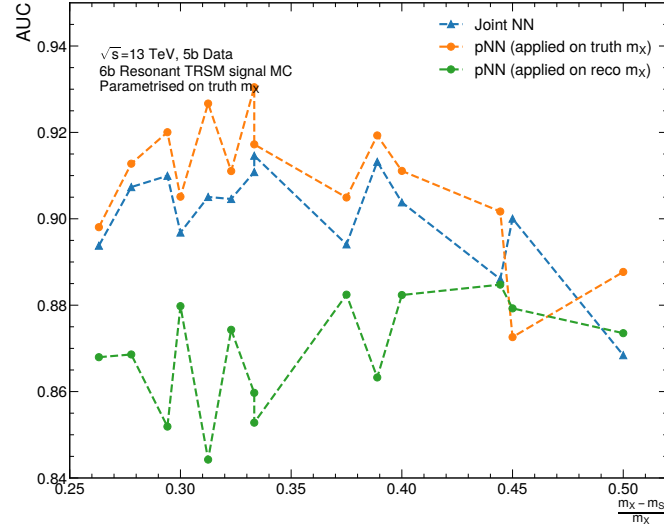


Figure 5.9: Area under the ROC curve (AUC) of the pNN and Joint NN evaluated on $6b$ resonant TRSM HHH signal samples and $5b$ data as background sample, as a function of the relative difference between the masses of the resonance particles $\frac{m_X - m_S}{m_X}$. The pNN is evaluated with data events assigned with the truth values of m_X in blue, and the reconstructed values in green. The performance is hindered when reconstructed values of m_X are assigned.

again a reduction in the performance but not as significant as the case where reconstructed parameter values are assigned. Shown in Figure 5.10, the pNN parametrised in this way out-performs the Joint NN across all true positive rates. This is due to the m_X values in the resonant TRSM HHH signal samples used as the physics parameters here being relatively close together, ranging only from 400 - 550 GeV. Hence the kinematics of the events with different m_X do not differ drastically, and the wrong assignment of the parameter value does not hinder the network performance significantly. However, if the kinematics differ significantly between the parameter values, the random assignment of them to events would be expected to have a much larger negative impact on the performance.

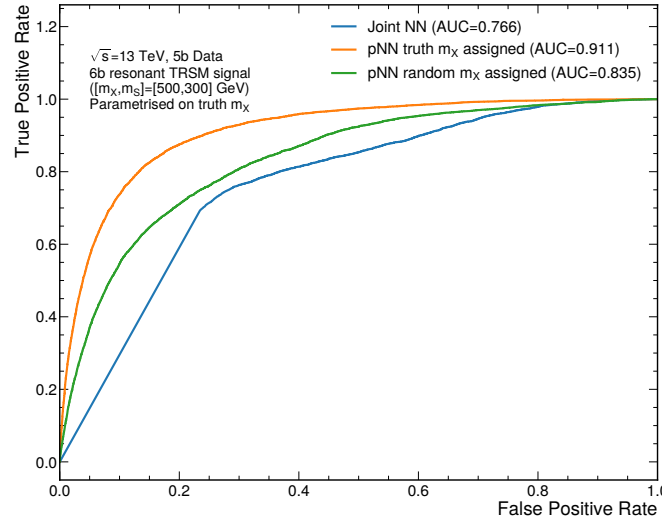


Figure 5.10: The ROC curve of the pNN and Joint NN evaluated on $6b$ resonant TRSM HHH signal samples and $5b$ data as background sample, as a function of the relative difference between the masses of the resonance particles $\frac{m_X - m_S}{m_X}$. The pNN is evaluated with data events assigned with the correct truth values of m_X in orange, and randomly assigned truth values in green. The pNN out-performs the Joint NN in both cases.

5.2.4 Conclusions

The advantages of using a parametrised neural network in the context of the $HHH \rightarrow 6b$ analysis is clear from the improved performance compared to the baseline joint NN. However, the suitable parametrisation scheme for the application to data is ambiguous, and three different approaches are explored in this study. Assigning the same truth m_X value as the physics parameter to all events for each signal mass hypothesis separately yields the best performance, but implies the most complex

analysis strategy, as a separate background estimate is needed for each mass hypothesis. Assigning the reconstructed m_X value of each event is not sensible due to the absence of the resonance particle in most data events, and therefore does not test any specific signal mass hypothesis. Randomly assigning the truth m_X values hinders the model performance, but not by a significant amount due to the similarity in kinematics between the events generated at different m_X values in the $HHH \rightarrow 6b$ signal samples. However, this approach would become less suitable if the kinematics differ significantly between the physics parameter values.

5.3 Kinematic Regression

Analyses often rely on constructing discriminating variables, or quantities of interest, that are sensitive to the target signal. These variables are usually constructed from physics objects, such as jets or leptons, selected by a set of criteria that target a signal topology. For example, for the resonant HHH production, the tri-Higgs invariant mass m_{HHH} corresponds to the mass of the heavy resonant particle X predicted by the TRSM model. With a $6b$ final state, m_{HHH} can be constructed by taking the invariant mass of the six b -jets that are produced from the decay of the three Higgs bosons in signal events. To further narrow down the phase space and improve signal sensitivity, one could attempt to assign the 6 jets into 3 pairs that correspond to three Higgs boson candidates, which then allows their individual masses to be reconstructed. In signal events, their masses would be close to the SM Higgs boson mass of 125 GeV. Since SM background events do not contain three genuine Higgs bosons, the masses of the reconstructed Higgs boson candidates would divert from this value. Hence in the (m_{H1}, m_{H2}, m_{H3}) volume, a Signal Region centred around (125, 125, 125) GeV can be defined to contain mostly signal events but few background events. This is a strategy akin to what was used in the Run-2 $HH \rightarrow 4b$ analysis, however there are limitations associated with this method:

- Selecting 6 jets passing a b -tagging working point significantly reduces signal efficiency (at a working point of 77% used in the Run-2 $HHH \rightarrow 6b$ analysis, roughly 22.8% of $6b$ events are selected).
- The decay of three Higgs bosons often results in more objects than 6 b -quarks, for example

from gluon radiation that leads to extra jets in the final state. Omitting them would impact the resolution of the reconstructed m_{HHH} .

- There are 15 unique ways of assigning 6 jets to 3 Higgs boson candidates, which is significantly more challenging than the $HH \rightarrow 4b$ case. This could lead to a lower pairing efficiency and therefore poorer reconstructed masses of the Higgs boson candidates.

An alternative approach of reconstructing quantities of interest is to directly regress them using machine learning. A mapping can be learnt between the observed quantities, such as jets, and a probability distribution of the quantities of interests, such as m_{HHH} .

$$f : \text{observed quantities} \rightarrow p(\text{quantities of interest}) \quad (5.1)$$

Machine learning algorithms are proficient at learning a function that performs such a mapping, such as the following:

$$f' : \text{observed quantities} \rightarrow \vec{\alpha} \quad (5.2)$$

where $\vec{\alpha}$ are the parameters of the probability distribution $p(\text{quantities of interest})$, such as the mean and standard deviation of a Gaussian distribution. To allow for the inclusion of all objects in an event, a Deep Sets neural network is chosen to learn this mapping, described in detail in Section 5.3.1. This approach bypasses the exploding combinatorics of jet pairing in multi-Higgs analyses that becomes increasingly challenging as the number of Higgs bosons increases.

5.3.1 Deep Sets Architecture

Physics events often have varied numbers of reconstructed objects, which can be represented by unordered, variable-sized sets. A Deep Sets network is a machine learning architecture capable of taking such sets as inputs, consisting of time distributed layers that iterate over the objects within each set to learn their features. In the $HHH \rightarrow 6b$ case, these sets are jets in each event. These learnt jet features are then summed and fed into a feedforward Deep Neural Network that provides the final outputs for each event. The schematic in Figure 5.11 summarises the architecture.

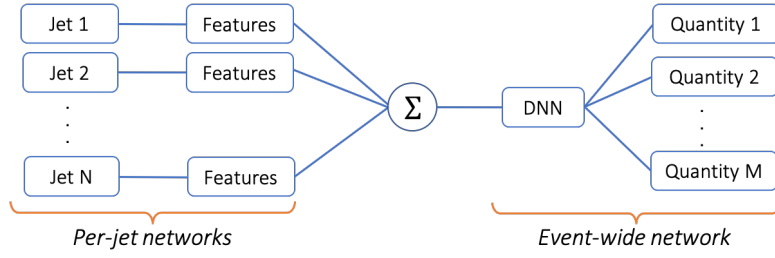


Figure 5.11: Schematic summarising a Deep Sets architecture for regressing M quantities of interest from events containing jets. The per-jet network iterates over all jets within the event. The sum of the learnt jet features becomes the input to the event-wide DNN that provides the final regression outputs.

5.3.2 Training

The input features to the Deep Sets network consist of events and their jet-level variables: jet p_T , η , ϕ and the b -tagging score quantile from the offline b -tagging algorithm for Run-2 DL1r. The target output of the network is the truth m_{HHH} for each event. For training, simulated SM HHH signal events passing the basic event pre-selections outlined in Section 4.3.3 are used. The SM HHH signal instead of the resonant HHH signals is used for its relatively flat distribution in m_{HHH} , which acts as a more general prior distribution of the regression algorithm. The resonant HHH signals have narrow and discrete peaks in their truth m_{HHH} distributions that correspond to various mass hypotheses of the heavy resonance particle X , which would bias the model towards these discrete masses. The network is implemented in Keras software, and trained with Tensorflow backend. The loss function minimises the mean squared error between the predicted and target values.

$$L_{MSE} = \frac{1}{n} \sum_{i=1}^n (y_{true,i} - y_{regressed,i})^2 \quad (5.3)$$

Details of the training parameters are summarised in Table 5.5.

5.3.3 Performance Comparison with Baseline

The regressed m_{HHH} values are compared to the truth values (response) in the validation dataset of the SM HHH signal sample. Their relative difference (resolution) is shown in Figure 5.12. The response is reasonably diagonal, but with a sizeable spread. This is also reflected in the resolution, where the relative differences from an approximate Gaussian with a mean of 0.05, but a

Parameter	Values
Learning rate	0.001
Batchsize	128
Number of jet layers	6
Number of event layers	6
Number of nodes in each jet layer	32
Number of nodes in each event layer	64
Number of epochs	100

Table 5.5: Hyperparameters of the Deep Sets Network for m_{HHH} regression.

relatively large standard deviation of 0.19. This indicates that the model is well-behaved, however the regression task is challenging likely due to the large number of objects in each event.

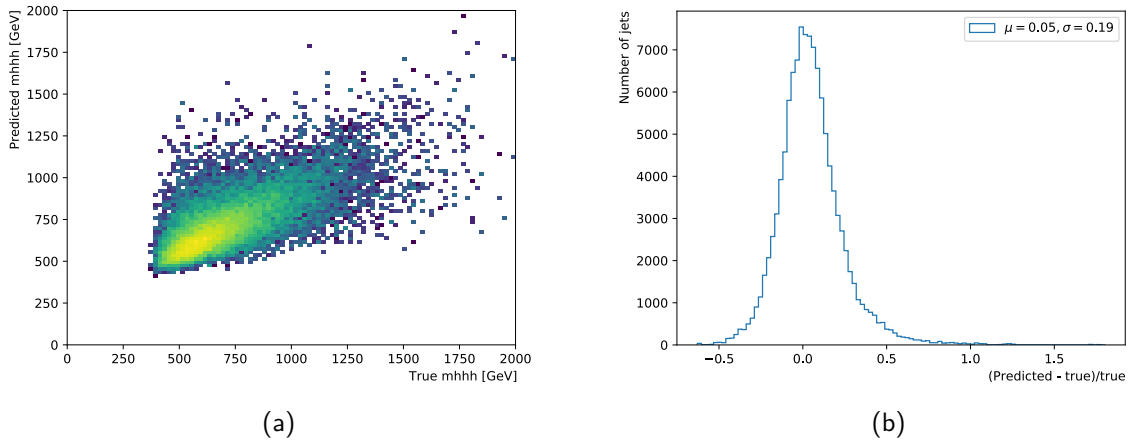


Figure 5.12: The response (a) and resolution (b) of the regressed compared to the truth m_{HHH} values evaluated on validation dataset of the SM HHH sample.

The trained model is applied to 2 representative resonant TRSM signal samples with $(m_X, m_S) = (450, 270)$ GeV and $(500, 300)$ GeV. The m_{HHH} values from regression are compared to the truth m_{HHH} values, m_{HHH} and m_{HHH125} reconstructed from taking the invariant mass of the three Higgs boson candidates reconstructed using a DNN-based pairing algorithm in Figure 5.13. The m_{HHH125} is calculated by rescaling the four-vector of each Higgs boson candidate to a mass of 125 GeV. The m_{HHH} values reconstructed from explicit jet pairing are closer to the truth values, while the regressed values consistently over-estimate. This is because the truth m_{HHH} distribution in the

SM HHH sample has a long tail in the higher values compared to the resonant TRSM signals. The model has therefore seen fewer examples of events with small truth m_{HHH} , and therefore has poorer performance in this region. The spread of the regressed m_{HHH} values is also larger compared to those reconstructed from explicit jet pairing. The spread of m_{HHH125} is even smaller due the ‘sharpening’ effect of the rescaling of the Higgs boson mass.

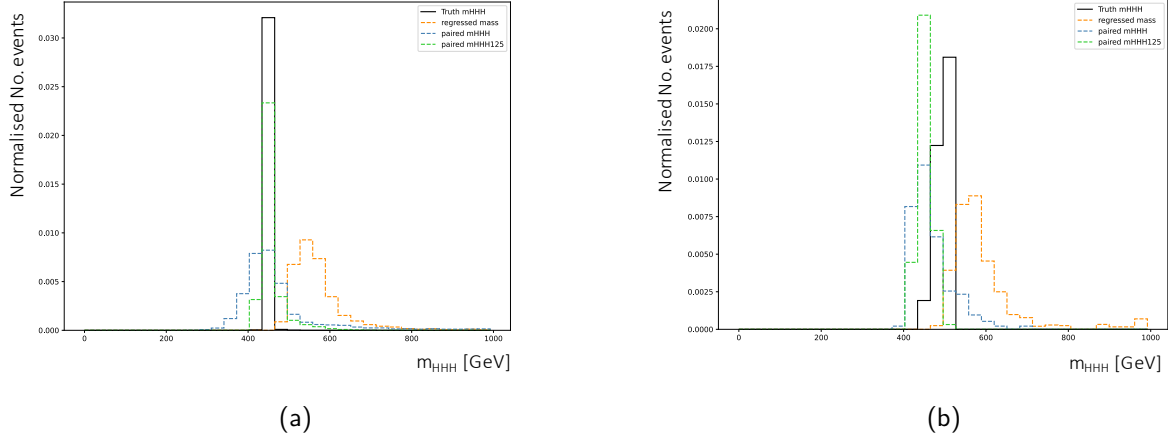


Figure 5.13: The truth m_{HHH} , the regressed m_{HHH} , and m_{HHH} and m_{HHH125} reconstructed from explicit jet pairing, evaluated on 2 representative resonant TRSM signals with $(m_X, m_S) = (450, 270)$ GeV (a) and $(500, 300)$ GeV (b).

5.3.4 Conclusion

The direct regression of m_{HHH} from jet-level variables using a Deep Sets network is explored. Although the performance does not exceed the conventional method of explicit jet pairing, the limitations of the method are understood for future usage. The main one lies in the truth m_{HHH} distribution used as the prior distribution in the training sample. To avoid introducing biases in the model, the ideal training sample would have a flat distribution in the target kinematic distribution which represents an ambiguous prior distribution. However, this is challenging for an analysis that searches for resonance particles in the target signals. For $HHH \rightarrow 6b$, the approach was to use the SM HHH signal sample for a relatively flat distribution in the truth m_{HHH} , however the SM signal could not represent the kinematics in the resonant TRMS signal well. This leads to the poor performance of the regression network when applied to the resonant signal samples that are unseen by the model. Another solution to obtaining a flat prior distribution is to combine the resonant

TRSM signal samples and resample the events to obtain a flat distribution in the truth m_{HHH} . However this would require large statistics in the signal samples to begin with, which is not available for this analysis. In conclusion, kinematic regression could be a powerful tool to construct kinematic quantities of interest from basic physics objects. The use of a Deep Sets architecture allows for the inclusion of all objects in an event instead of more stringent event selections that would limit signal efficiencies and sensitivity. While challenging for resonance searches, this approach could potentially be useful for analyses that have a large number of combinatorics, or missing energy in the final state.

6 | Dataset-wide Graph Neural Networks for Searches of New Physics

We should all do a bit more of hanging around,
lucid dreaming.

— Anna Mullin

6.1 Introduction

The complex and high-dimensional nature of data in particle physics has led to the rapid adoption of machine learning (ML) methods within the field. At the ATLAS experiment, ML methods have become commonplace in physics object reconstruction, calibration and physics analyses. The numerous ML-related projects featured in this thesis are only a glimpse into the extensive activities across a broad range of the physics topics.

In many scientific fields, graphs provide intuitive representations of information structures, such as social network and gene structure. In particle physics, they offer a natural way of representing measurements of physical quantities, such as detector hit signals associated with the trajectory of a charged particle, or physical objects produced in a proton-proton collision event. In a graph, the individual objects are represented as *nodes*, and the connections between them, referred to as *edges*, allow for the flow of information between nodes, capturing additional dynamics and relations. Graph Neural Networks (GNNs) [105] utilise this additional information contained within graph representations, and are becoming increasingly popular within particle physics. Their performance often surpasses conventional ML methods such as Boosted Decision Trees (BDTs) or Deep Neural Networks (DNNs) which treat objects as independent and identically distributed. Noteworthy GNN developments include the b -tagging algorithm GN1 [65] which utilises graph representations of jets to capture their substructures and the correlation between tracks within the jet cone. The observation of four-top-quark production at the ATLAS experiment in 2023 [106] used a GNN to separate $t\bar{t}t\bar{t}$ signal from a variety of complex background signatures. In general, these GNNs perform *graph*

classification, exploiting the topological differences between graphs. For example in GN1, each jet is represented by a graph consisting of nodes that correspond to the tracks within the jet cone, and graph classification is performed to identify the quark flavour of each jet. Here, the nodes are characterised by the kinematic features of the individual tracks and their associated jet. In the four-top-quark analysis, the GNN presents each event as a graph, where each node corresponds to a reconstructed physics object in the event. In this case, graph classification is carried out to separate signal from background events. Here, the nodes are characterised by the kinematic features of the individual objects. Both applications utilise fully connected graphs, where each node is connected to all other nodes. The approach presented in this study however, considers a single graph representation of an entire dataset that consists of both signal and background events. Here, each node is an event in the dataset and the edges are made based on the kinematic *similarity* between events. This naturally leads to topological features within the graph representation that are unique to individual physics processes, which can be exploited by a GNN to perform *node classification*. For the discussions on the graph representation used by such GNNs, nodes and events are used interchangeably in the following sections.

New physics signatures from BSM models are challenging to detect as they often produce events that are extremely similar to ones associated with SM processes. Differences in event correlation and structural features between BSM signal and SM background events can be used as additional information for their separation. For example, a BSM model that hypothesises a heavy resonance particle decaying into SM particles would result in events with a peak in the reconstructed invariant mass of the decay products, which corresponds to the mass of the heavy resonance particle. However, the reconstructed invariant mass of the same decay products from SM processes usually would not result in a sharp peak in the distribution. Taking this invariant mass variable as a one-dimensional domain, the BSM signal events are more *clustered* around the mass of the heavy resonance particle compared to the SM background events. Extending this domain to higher dimensions, for example by adding more reconstructed kinematic variables, more detailed topologies can be revealed. A graph representation that is dataset-wide can capture these features by considering the relationships between all events in a dataset, hence adding to the discriminating power between nodes.

The work presented in this thesis is strongly inspired by the approach explored in Reference [107], where a dataset-wide graph representation is constructed to extract discriminating network metrics to improve BSM signal and SM background discrimination. These metrics are used to capture the graph topological features in a high-dimensional space, such as *degree centrality*, defined as the number of nodes connected to a given node, and *betweenness centrality*, defined as the fraction of shortest paths between pairs of randomly chosen nodes that pass through a given node. They are node-splitting invariant, meaning that the calculations account for the effective event yields expected in a collider experiment which can be encoded in node weights. Excellent separation between the BSM signal and the SM background events is observed in these network metric distributions, proving them to be as powerful as conventional kinematic variables. Instead of explicitly calculating these network metrics, this work captures the graph topology implicitly by embedding the graph representation in a GNN. The difference in connectivity between different classes of events are revealed through the convolution of their features during the GNN training, and thus exploited to improve classification.

The theoretical background of GNNs and the different types of graph convolution operation are discussed in Section 6.2. A case study on a BSM leptoquark signature is presented in Section 6.3, which includes the details of its production process and the simulation of its events used as training and validation samples in this study. Details about the GNN model training and implementation are described in Section 6.5, in particular the challenge of model scalability is discussed in Section 6.4. The performance of an array of GNN models experimenting with different node features, graph construction and graph convolutions are documented in Section 6.6. For each class of node features, a DNN is trained as a baseline model for comparison. Finally, conclusions and outlooks on this effort are summarised in Section 6.7 and Section 6.8.

6.2 Theoretical Background

The notation and construction of a graph are first introduced in Section 6.2.1. The mathematical expressions of the graph convolution performed in three types of GNNs are outlined in Section 6.2.4. For this study, two types of GNNs with different convolution operations are explored: the Graph Convolution Network (GCN) [108] and the Higher-order Graph Neural Network (GraphConv) [109]. They are implemented in the machine learning library PyTorch Geometric [110] which is used to perform all GNN trainings for this study. The mathematical formalism for each will be summarised in Section 6.2.4.

6.2.1 Graph Construction

A graph $G = (N, E)$ consists of a set of nodes N connected by a set of edges $E \subseteq \{\{i, j\} \subseteq N \mid i \neq j\}$. Each node in the graph is associated with a vector of node features, where the feature k of node i is denoted x_{ik} . For a graph with n nodes, their features together form the node feature matrix $\mathbf{X}$ of dimension $(n \times k)$, which become the input features to a GNN. The edges between nodes can be represented by an *adjacency matrix* $\mathbf{A} = (a_{ij})_{i,j \in N}$ of dimension $(n \times n)$, where

$$a_{ij} = \begin{cases} 1 & \text{if } i, j \in E, \\ 0 & \text{otherwise,} \end{cases} \quad (6.1)$$

It is also important to include each node's connection to itself (self-connection), in order to preserve its original features during the convolution, especially in the case where a node is not connected to other nodes. To include the self-connection, an *extended adjacency matrix* can be defined as $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$, where its elements are $\tilde{a}_{ij} = a_{ij} + \delta_{ij}$. A simple example of a graph consisting of 3 nodes in a domain with 2 features and its corresponding extended adjacency matrix are shown in Figure 6.1.

For a given node u , the set of nodes immediately connected to it are its *nearest neighbours*, defined

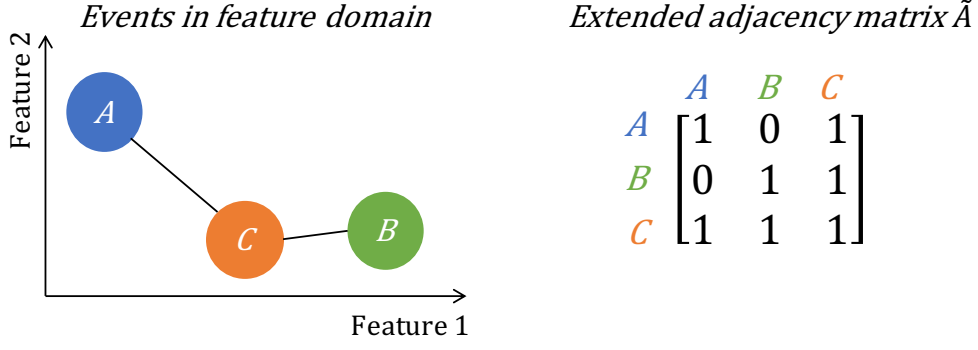


Figure 6.1: An example of a simple graph consisting of 3 nodes in a domain constructed from 2 features. The edges between the nodes are encoded in the extended adjacency matrix, where an edge between two nodes is represented by a value of 1, otherwise by a value of 0. Self-connections are accounted for by setting the diagonal elements to 1.

as

$$N_u = \{v \in N \mid \tilde{a}_{uv} = 1\} \quad (6.2)$$

which form its *neighbourhood*. The set of edges E in the graph is determined by the distance $d_{i,j}$ between each pair of nodes $i, j \in N$ in a chosen domain, calculated using a chosen distance metric. The domain is constructed from a set of features associated with each node, forming a k -dimensional space where k is the number of features. Note that it is not necessarily the same set of node features used as input features for the GNN training downstream. Within it, a distance metric can be defined to quantify the similarity between pairs of events. The choice of distance metric depends on the targeted signature and the chosen kinematic variables that construct the domain. Three distance metrics are explored for this study, which are detailed in Section 6.2.2.

The adjacency matrix is defined as

$$a_{ij} = \begin{cases} 1 & \text{if } d_{i,j} < LL, \\ 0 & \text{otherwise,} \end{cases} \quad (6.3)$$

where LL is referred to as the *linking length*. This length can be chosen using various metrics depending on the use for the graph representation. For example, it can be determined by the edge fraction in the graph, defined as the portion of edges in all possible edges. This has important

impacts on the memory usage of the GNN training downstream, as a higher edge fraction requires a higher number of convolution calculations.

6.2.2 Distance Metrics

The similarity between a pair of events u, v can be quantified using

- Euclidean distance: $d_{u,v}^{Euclidean} = \sqrt{\sum_k (u_k - v_k)^2}$
- Cosine distance: $d_{u,v}^{cosine} = 1 - \frac{u \cdot v}{\sqrt{u \cdot u} \sqrt{v \cdot v}}$
- Earth Mover's Distance (EMD) [111], [112]: the minimum energy required to transform one event into another in the 4-momentum space

where k is the number of kinematic features associated with each event that constructs the domain. The values of these distances are sensitive to the scale of the kinematic features, whereby the ones with larger values contribute more to the distance calculation. In order to eliminate this effect, the kinematic variables of all events are standardised to have a mean of zero and a standard deviation of one before distance calculation.

The Euclidean distance quantifies the absolute magnitude of the separation between pairs of events within a domain, whereas the cosine distance quantifies the angular separation between vectors formed from the origin to the two events in a domain. The EMD is motivated by its physical interpretability, which warrants a detailed mention. It represents the minimum amount of energy required to transform an event u to another event v . This energy is broken down into the components required to move a physical object i in event u to an object j in event v , such that

$$\sum_{i,j} f_{ij} = \min \left(\sum_i E_i, \sum_j E_j \right) \quad (6.4)$$

where E_i and E_j are the energies of the physical objects i and j . The f_{ij} is also referred to as the "energy flow" from one physics object to another [112]. Aggregating over all the energy flows, the

EMD between events u and v is defined as

$$d_{\text{EMD}}(u, v) = \min_{f_{ij}} \left(\sum_{i,j} f_{ij} \frac{\theta_{ij}}{\alpha} \right) + \left| \sum_i E_i - \sum_j E_j \right| \quad (6.5)$$

where i and j index the physical objects in event u and v separately, θ_{ij} is the angular separation between objects, and α controls the relative importance between the first and second term in the equation. The first term minimises the amount of energy required to move the particles in event u to event v in the four-momentum space. The second term accounts for the absolute difference in energy between the two events, such that the transformation follows the conservation of energy.

The scale of the distances can vary depending on the metric used, therefore a separate linking length is calculated for each distance metric. In this study, similar events, that is events with distances smaller than a linking length are connected, forming a "friend graph". Inversely, an "enemy graph" can also be constructed by connecting events distances larger than a linking length, which could be interesting for certain physics processes but was not explored in this study.

6.2.3 Graph Neural Network

A GNN is a neural network model that learns patterns from the graph representation of a dataset by aggregating information between the nodes. It takes a set of attributes associated with the nodes as input features and an adjacency matrix which encodes the connections between them. Through every hidden layer of the network, a convolution is performed that aggregates the features of each node with that of its neighbours. The output of the l^{th} hidden layer can be expressed as a function of the extended adjacency matrix $\tilde{\mathbf{A}}$ and outputs from the previous hidden layer ($l - 1$)

$$H^{(l)} = f(H^{(l-1)}, \tilde{\mathbf{A}}) \quad (6.6)$$

where $H^{(0)} = \mathbf{X}$ is the input layer. The expression for function f depends on the type of graph convolution being performed, two of which are outlined in Section 6.2.4. The final layer of the GNN produces $H^{(L)} = Z$ where Z are the model outputs per data point. In the case of a binary classifier

for signal and background classes, $Z \in [0, 1]$ corresponds to the probability that an event belongs to the signal class.

A simple convolution function for the l^{th} layer can be

$$f(H^{(l)}, \tilde{\mathbf{A}}) = \sigma(\tilde{\mathbf{A}}H^{(l-1)}\mathbf{W}^{(l)} + \mathbf{b}^{(l)}) \quad (6.7)$$

where σ is a non-linear activation function such as the sigmoid function¹. $\mathbf{W}^{(l)}$ is a matrix of trainable weights and $\mathbf{b}^{(l)}$ a matrix of trainable biases in the l^{th} hidden layer. Their elements are updated during each iteration of the training. The matrix multiplication and the dimensions of the matrices are summarised in Figure 6.2.

$$\begin{array}{c}
 \begin{array}{cccc}
 & \tilde{\mathbf{A}} & & H^{(l-1)} & & \mathbf{W}^{(l)} & & \mathbf{b}^{(l)} \\
 H^{(l)} = & \left\{ \begin{bmatrix} 1 & 0 & 1 & & \\ 0 & 1 & 1 & \dots & \\ 1 & 1 & 1 & & \\ & \vdots & & \ddots & \\ & & & & 1 \end{bmatrix} \right. & \times & \begin{bmatrix} x_{11} & \dots & x_{1h'} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nh'} \end{bmatrix} & \left. \right\} & \times & \begin{bmatrix} w_{11} & \dots & w_{1h} \\ \vdots & \ddots & \vdots \\ w_{h'1} & \dots & w_{h'h} \end{bmatrix} & + & \begin{bmatrix} b_{11} & \dots & b_{1h} \\ \vdots & \ddots & \vdots \\ b_{k1} & \dots & b_{kh} \end{bmatrix} \\
 & \underbrace{\begin{array}{c} [n \times n] \\ \times \\ [n \times h'] \end{array}}_{[n \times h']} & & & \times & [h' \times h] & + & [n \times h] \\
 & & & & \underbrace{\begin{array}{c} [n \times h'] \\ \times \\ [h' \times h] \end{array}}_{[n \times h]} & + & [n \times h] \\
 & & & & & \underbrace{\begin{array}{c} [n \times h] \\ + \\ [n \times h] \end{array}}_{[n \times h]}
 \end{array}
 \end{array}$$

Figure 6.2: The matrix multiplication of a simple graph convolution function in Equation 6.7 for the l^{th} hidden layer of a GNN, annotated with the dimension of the matrices in each step of the multiplication. The extended adjacency matrix is denoted $\tilde{\mathbf{A}}$ with dimension $(n \times n)$ where n is the total number of nodes in the graph. The output of the $(l-1)^{th}$ hidden layer of the GNN is a $(n \times h')$ matrix $H^{(l-1)}$, where h' is the number of hidden nodes in the previous hidden layer. The trainable parameters of the l^{th} hidden layer are denoted $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$. The output of this convolution operation yields a matrix of dimension $(n \times h)$, where h is the number of hidden nodes in the l^{th} hidden layer.

The multiplication with $\tilde{\mathbf{A}}$ performs a convolution of the outputs from the previous layer, where the features of each node are summed with the input features of its nearest neighbours. In other words,

¹The sigmoid function is expressed as $f_{sigmoid}(x) = \frac{1}{1+e^{-x}}$

each target node receives updates from all its nearest neighbours in one "message passing" step, illustrated in Figure 6.3, in the first hidden layer of the GNN. In the second hidden layer, a target node receives updates from all of its nearest neighbours, but also their nearest neighbours. This convolution is performed as many times as there are hidden layers in the network, hence convoluting over increasing numbers of connected nodes in the graph.

Additional features can be encoded in the edges E as *edge weights*. They are incorporated in the message passing step by weighting the update a target node i receives from a nearest neighbour j by the edge weight $e_{j,i}$ between them. In this study, two features are represented in the edge weights: the similarity of a pair of events, and the effective yield of each event. The former is captured by the inverse distance between pairs of events, such that the message passing between those with smaller distances are given more importance. The latter is captured by the weights assigned to each event during MC simulation w_{event}^{MC} , representing its effective yield accounting for the theory predicted cross-section of the underlying physics process, detector luminosity and efficiencies. They are necessary as it is often impractical to simulate the number of events expected in the detector given these parameters. For example, a common SM process would result in an unfeasible amount of events to simulate, but a rare or hypothetical BSM signal would often result in a small amount of statistics incapable of capturing the full kinematic spectrum. As events are not treated as independent nodes in a graph, the topologies are dependent on the raw number of nodes present. The w_{event}^{MC} effectively represents the expected number of events in the detector, and utilises the node-splitting-invariant (n.s.i.) property of the convolution performed during message passing. Discussed in more details in [107], this property entails that 2 fully connected nodes with identical features can be represented by a singular node with a weight of 2. The w_{event}^{MC} play such a role and are therefore essential in ensuring the GNN trained on simulated events is robust when applied to recorded data.

Combining both features above, the expression for the edge weight between a target node i and

each of its nearest neighbours j is therefore

$$e_{j,i} = f_{\text{Minmax}}\left(\frac{1}{\tilde{d}_{i,j}}\right) \times w_{event}^{MC} \quad (6.8)$$

where $\tilde{d}_{i,j}$ is the distance between the neighbouring nodes with $\tilde{d}_{i,j} = 1$ when $i = j$ to account for self-connections. The Minmax normalisation function² rescales the inverse distances to a fixed range between 0 and 1 which prevents extreme values in the edge weights that can lead to instability in the model training.

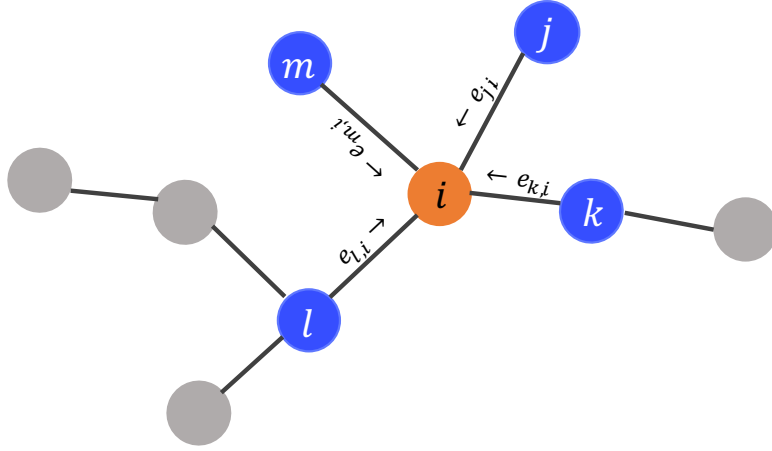


Figure 6.3: A schematic of the message passing of an example graph, where the features of a target node i in orange receive updates from all its nearest neighbours (in blue). The edge weights between each nearest neighbour and the target node are also labelled.

6.2.4 Graph Convolution Functions

Different convolution operations can be performed to aggregate features between connected nodes. The node features are organised into a matrix $\mathbf{X}$ of dimension $(n \times k)$, where n is the total number of events in the dataset, and k is the number of kinematic features for each event. As they are propagated through a multi-layer GNN, they are convoluted due to the connectivity between nodes that is encoded in the adjacency matrix $\tilde{\mathbf{A}}$. Two convolution functions, namely the Graph Convolution Network and the Higher-order Graph Neural Network are explored in this study.

²The Minmax function of a distribution x is expressed as $f_{\text{Minmax}}(x) = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$.

Graph Convolution Network (GCN) In a GCN network [108], the input features $\mathbf{X}$ are convoluted in the first hidden layer as the following

$$\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X} \quad (6.9)$$

where the $\tilde{\mathbf{D}}$ is the *degree centrality matrix*, a diagonal matrix where the elements correspond to the number of nearest neighbours connected to each node, such that $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$. The $\tilde{\mathbf{D}}^{1/2}$ represents the square root of its diagonal elements. The operation $\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2}$ effectively normalises the adjacency matrix by the degree centrality of each node, such that all rows in the normalised matrix sum to unity. This prevents the scale of the input features to be changed by simply performing a summing convolution. It also eliminates the "celebrity effect", where the kinematic features of nodes with larger number of neighbours are less unique as they are convoluted with more nodes, making them less unique. $\mathbf{X}'$ is also referred to as the once convoluted kinematics of events in the dataset. Subsequently, it is treated with a non-linear activation function in the first hidden layer along with its trainable parameters, and becomes the inputs to the second hidden layer. The convolution function of each hidden layer in the GCN is therefore

$$H^{(l)} = \sigma(\tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} H^{(l-1)} \mathbf{W}^{(l)}) \quad (6.10)$$

where σ is a non-linear activation function. With the inclusion of edge weights, the convoluted kinematics of a given node i become

$$\mathbf{x}'_i = \sum_{j \in \text{Neighbourhood}(i)} \frac{e_{j,i}}{\sqrt{\hat{d}_j \hat{d}_i}} \mathbf{x}_j \quad (6.11)$$

where $\hat{d}_i = 1 + \sum_{j \in \text{Neighbourhood}(i)} e_{j,i}$.

Higher-order Graph Neural Network (GraphConv) The graph convolution of each hidden layer in a GraphConv network is based on the 1-dimensional Weisfeiler-Leman (WL) algorithm [113], which is commonly used to determine the isomorphism between two graphs by iteratively assigning locally aggregated features to nodes until they converge. The isomorphism testing compares the assigned

colourings between two graphs. The assignment of aggregated features is analogous to message passing in a GNN, where the node features are repeatedly updated by the features of their nearest neighbours through the hidden layers. The GraphConv network generalises the 1-dimensional WL algorithms to k dimensions, such that aggregated features are assigned to k tuple of nodes instead of singular nodes. This allows higher-order features of subgraphs that are not represented at the node-level to be learned. The convoluted feature of a given node i in layer l can be expressed as

$$f^{(l)}(i) = \sigma \left(f^{(l-1)}(i) \cdot W_1^{(l)} + \sum_{j \in \text{Neighbourhood}(i)} (e_{j,i} \cdot f^{(l-1)}(j)) \cdot W_2^{(l)} \right) \quad (6.12)$$

where the target node i receives updates from each of its nearest neighbour j . Here, $W_1^{(l)}$ and $W_2^{(l)}$ are trainable parameters matrices. A new feature which aggregates over the neighbourhood features from Equation 6.12 and the node representation from layer $(l - 1)$ can be expressed as:

$$f^{(l)}(i) = f_{merge}^{W_1} \left(H^{(l-1)}(i), f_{aggr}^{W_2}(H^{(l-1)}(j) \mid j \in \text{Neighbourhood}(i)) \right) \quad (6.13)$$

where $f_{aggr}^{W_2}$ aggregates over the features of a neighbourhood of nodes, and $f_{merge}^{W_1}$ merges the the aggregated features with ones from the $(l - 1)^{th}$ iteration. The dimension k is equal to the number of nodes within the subgraph of a neighbourhood where the features are learned, illustrated in Figure 6.4.

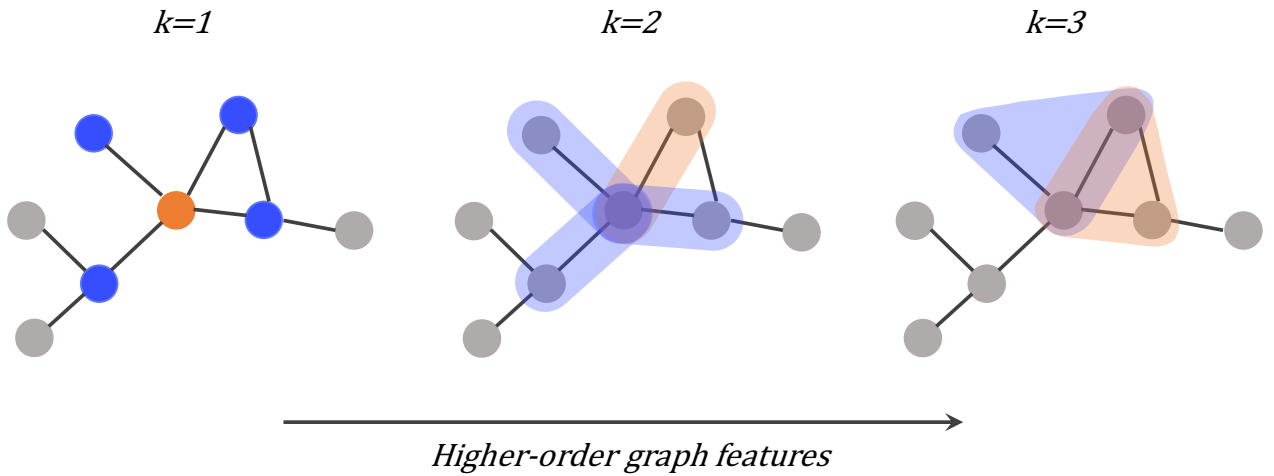


Figure 6.4: Schematics of higher-order graph features within a graph aggregated through each hidden layer of a GraphConv network. In each layer, the features of the k -element subgraph are learned.

6.3 Case Study: Searching for Leptoquarks

To explore the potential of this dataset-wide approach in the search for new physics over the dominant SM background processes, a BSM leptoquark (LQ) signal is used as a case study. Leptoquarks [114] are hypothetical particles that provide direct coupling between leptons and quarks, thus bridging the gap between these two sectors. Motivated by a wide range of theories, such as the unification of particle interactions and certain classes of supersymmetry, they predict unique kinematic properties.

For this study, the chosen model predicts a LQ as a vector boson U_1 that carries colour charge, weak isospin charges 3 and 1, and an electric charge of $2/3$. Each LQ then subsequently decays into a top quark and tau-neutrino ν_τ at 100% branching fraction. The decay of the top quark produces a b -quark and a W boson, whose leptonic decay is targeted for more efficient triggering in the detector due to the leptons. The final state contains a pair of reconstructed b -jets, a pair of oppositely-charged leptons, and missing energy E_T^{miss} attributed to the neutrinos produced from the decay of the LQ and the W boson, as shown in Figure 6.5. SM processes such as the production of a single top-quark (single-top) and $t\bar{t}$ produce similar final states, constituting the major backgrounds for this signal. Whilst scalar LQ models with this coupling and branching fraction setup have been excluded by the ATLAS experiment below 1.2 TeV, and U_1 with a branching fraction set to 50% excluded up to 1.84 TeV [115], no direct constraints have been set on this model. The details on the simulation of the LQ signal and the SM background processes are detailed in Section 6.3.1.

6.3.1 Simulated Events and Event Selection

Monte Carlo simulated events are produced to represent the pair-produced LQ signal and dominant $t\bar{t}$ and single top production background processes. Events are generated at a centre-of-mass energy of 13.6 TeV using MADGRAPH_AMC@NLO 2.9.21 [116], interfaced with PYTHIA 8.311 [117] for the parton shower and hadronisation. The events are generated at leading order in the Matrix Element using the NNPDF2.3LO PDF set [87], with the top quark mass set to 172.5 GeV. To ensure the extreme tails of the kinematic distributions are adequately sampled, events are generated in orthogonal slices of E_T^{miss} value. More events are generated for the slices in these tail

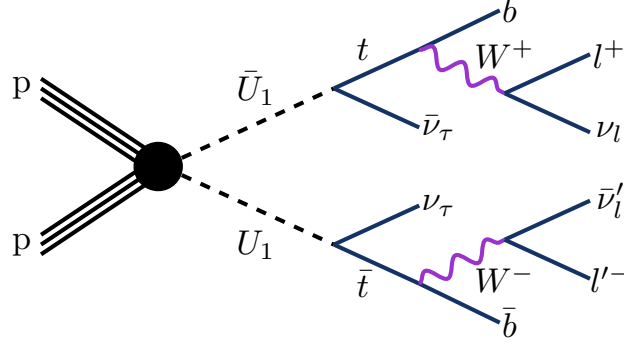


Figure 6.5: Feynman diagram of the production of a pair of leptoquarks $\bar{U}_1$, which subsequently decay into a pair of top-quarks and τ -neutrinos. Each top-quark decays into a bottom-quark and a W -boson, which then decay leptonically into a charged lepton and neutrino. The resultant final state consists of a pair of bottom quarks from the top-quark decays, a pair of oppositely-charged leptons from the W -boson decays, and E_T^{miss} corresponding to the four neutrinos from the leptoquark and W -boson decays. The diagram is made by Dr. Holly Pacey in preparation of the paper publication of this study.

regions, and their contribution to the overall sample are weighted down to represent their realistic effective yields produced in the detector. Simulation of the ATLAS detector response is performed using DELPHES 3.5.0 [118] using the default ATLAS detector card. Jets are reconstructed using the anti- k_T algorithm [49] with a radius parameter of $R = 0.4$ using the FastJet package [119].

For the LQ signals the "VectorLQ" Feynrules model is used from the LQ Toolbox [114], [120]. The coupling strength to a top and tau-neutrino g_{tL} is set to 1 whilst all other couplings are set to 0. As such the branching fraction for $U_1 \rightarrow t\nu_\tau$ decay is 1, to target the process of interest. The "Yang-Mill" scenario is used, where a 4-point coupling between two U_1 s and two gluons is introduced, enhancing the LQ pair-production cross section. The mass of the LQ is set to 1 TeV by default. The process generated produces two U_1 LQs which both decay into a top quark and neutrino. The top quark is decayed leptonically (including τ leptons) in MadGraph.

For $t\bar{t}$ and single-top samples the 5-flavour scheme is used, and the top quark is decayed leptonically (including τ leptons) in MadGraph. The $t\bar{t}$ samples are normalised to the cross-section calculated at Next-to-Next-to-Leading-Order (NNLO) in perturbative QCD and with Next-to-Leading-Log order soft-gluon re-summation [121]. The single top samples are normalised to cross-sections calculated at NNLO in perturbative QCD for s- and t-channel production [122], and at Next-to-Leading order

in perturbative QCD with Next-to-Next-to-Leading-Log order soft-gluon re-summation for tW production [123].

The decays of the pair of LQs and the semi-leptonic decays of the W -bosons produce multiple neutrinos in the final state, reconstructed as a relatively large amount of E_T^{miss} compared to the SM background processes. Events are required to have exactly two electrons, muons or one electron and one muon, that are oppositely-charged and have a $p_T > 20$ GeV and $|\eta| < 2.5$. Exactly 2 b -jets are also required, each with a minimum p_T of 20 GeV and $|\eta| < 2.5$. In order to select more signal-like events, only events with $E_T^{\text{miss}} > 200$ GeV are selected for the training and validation of the GNNs in this study.

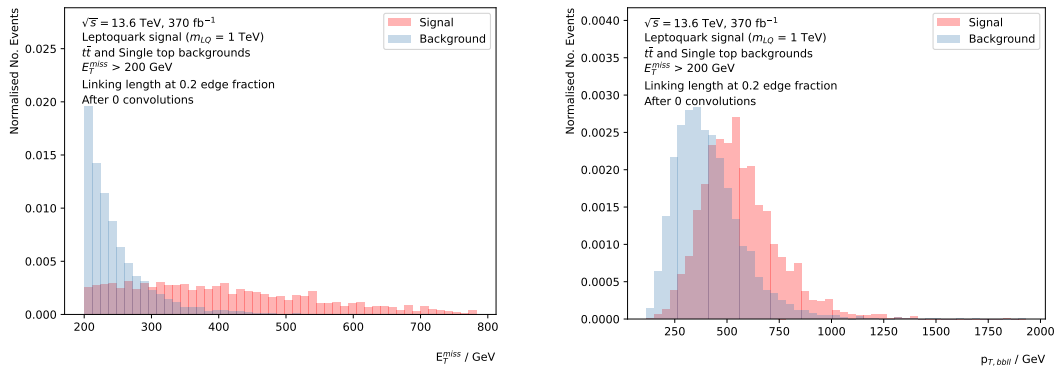
6.3.2 Event Similarity in Kinematic Space

The similarity between physical events are quantified as the distance between them in a chosen domain. To accentuate the topological differences between signal and background events, it is beneficial to construct a domain using a set of variables that maximises the disparity in the distance distributions between them. As physicists, it is possible to feature engineer the input variables using theoretical or phenomenological knowledge about the physics processes and provide a set of high-level kinematic variables. However, this comes at a cost of the method's ability to generalise to different kinds of physics processes and final states. Using a set of low-level kinematic variables such as the four-momentum of particles in the final state, is more general but less discriminating. The graph domain constructed from both types of variables are discussed in the context of the LQ case study.

For the targeted LQ signal, the large number of invisible particles in the final state leads to larger reconstructed E_T^{miss} on average compared to the SM background processes, making it a natural choice of discriminating kinematic variable. Due to the production of neutrinos from the leptonic decay of the W bosons, the masses of leptons can only be reconstructed in the transverse direction. They tend to be larger for LQ signal events, as the t -quark produced from a heavy LQ is more boosted, producing more boosted leptons in the subsequent decays. This results in a larger sum

of the p_T of the 2 b -jets and leptons, making it a discriminating kinematic variable. One could also include the approximate transverse mass of the two W bosons reconstructed using E_T^{miss} and the two leptons $m_{T,l1}$, $m_{T,l2}$ from the leptonic decay of the W bosons. Due to the presence of multiple invisible particles in the final state in the LQ signature, the visible physics objects, such as the reconstructed b -jets and the leptons, are expected to have larger angular separation between them. Kinematic variables related to these angular separations are therefore discriminating, such as minimum separation in ϕ between the leading lepton and E_T^{miss} . Together the 5 kinematic variables: E_T^{miss} , $p_{T,bbl}$, $m_{T,l1}$, $m_{T,l2}$, and $\text{Min}\Delta\phi_{E_T^{\text{miss}},l1}$, form a 5-dimensional high-level kinematic space, where the distances between the events within the dataset can be calculated.

The distributions of these kinematic variables between a LQ signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds are shown in Figure 6.6. The pair-wise Euclidean and cosine distances between pairs of signal events (sig-sig), pairs of background events (bkg-bkg) and pairs of signal and background events (sig-bkg) within this kinematic space are shown in Figure 6.7. In this kinematic space, the background events are more clustered together in terms of magnitude, characterised by a larger proportion of background pairs with small Euclidean distances. However they are oriented more sparsely in direction, characterised by a larger proportion of background pairs with large cosine distances. The separations between sig-sig, sig-bkg and bkg-bkg distributions are quantified using the Kullback-Leibler (KL) divergence³, which measures the similarity between two distributions.



³The KL divergence between distributions $p(x)$ and $q(x)$ is calculated as $D_{KL} = \sum_x p(x) \ln \frac{p(x)}{q(x)}$.

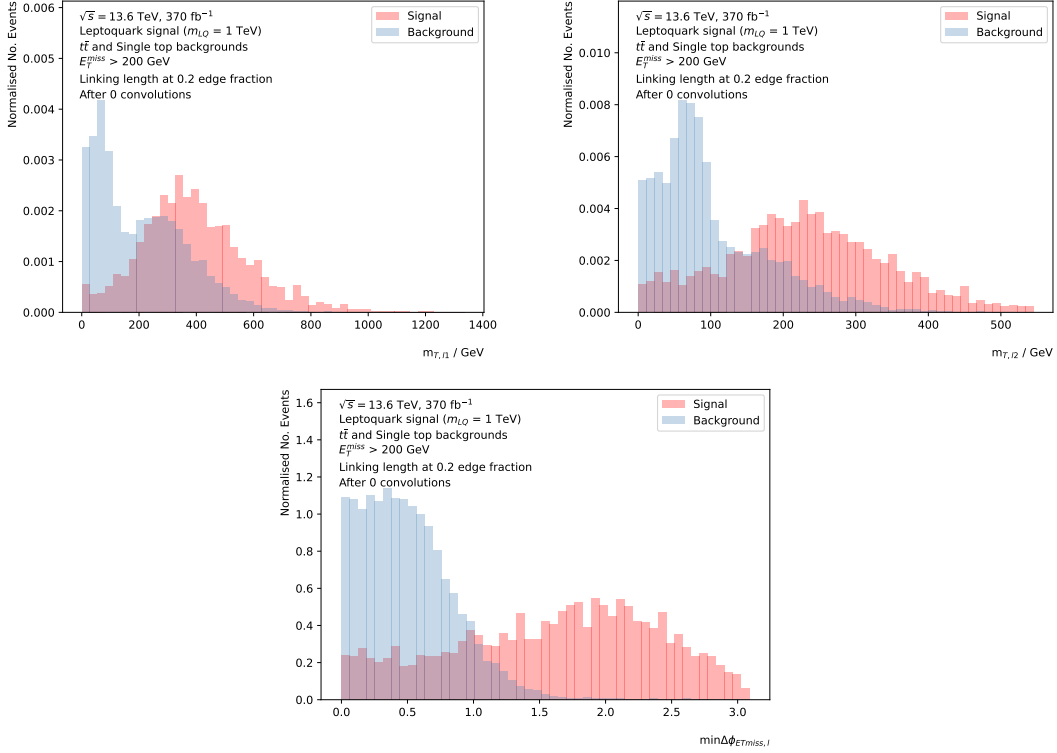


Figure 6.6: Distributions of the the high-level kinematic variables, E_T^{miss} , $p_{T,bbl}$, $m_{T,l1}$, $m_{T,l2}$, and $\text{Min}\Delta\phi_{E_T^{\text{miss}},l1}$ between LQ signal with $m_{LQ} = 1$ TeV and SM $t\bar{t}$ and single top backgrounds. They form the 5-dimensional kinematic space where the distances between events can be calculated. All histograms are normalised to have unit area to allow for shape-only comparisons between the distributions. These variables are chosen to maximise the topological differences between LQ signals and SM background using physics knowledge.

Low-level kinematic variables are more general as they do not require specific physics knowledge about the targeted signal, but are also less discriminating. For example, they could be the p_T , η and ϕ of each reconstructed decay product of the LQ signal, namely the two b -jets and leptons, as well as the E_T^{miss} from the presence of neutrinos and its distribution in ϕ within the detector. The distributions of these kinematic variables are shown in Figure 6.8. Together, they form a 14-dimensional low-level kinematic space, and the Euclidean and cosine distances between pairs of events are shown in Figure 6.9. The separations between the distance distributions are not as large as ones in the kinematic space constructed from high-level kinematic variables, evident in the KL divergence values between the same and different-class distance distributions. This is expected as the low-level kinematics, albeit more general, are not chosen with any expert knowledge, and are therefore less discriminating.

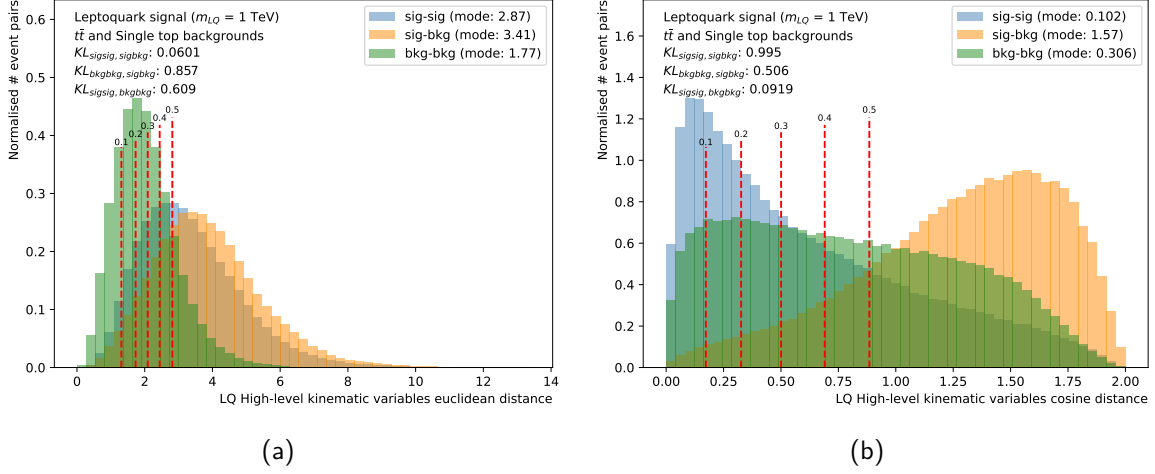
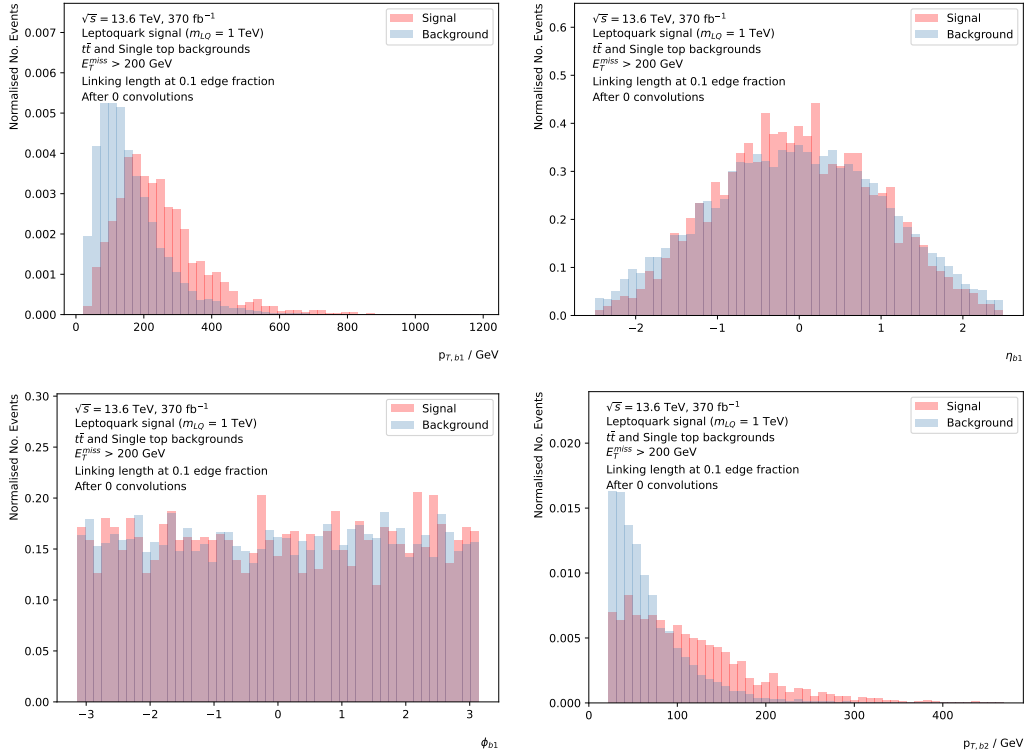
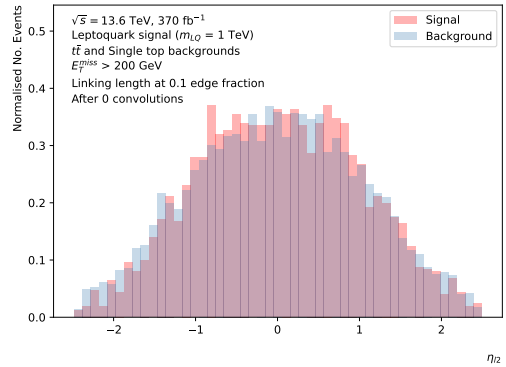
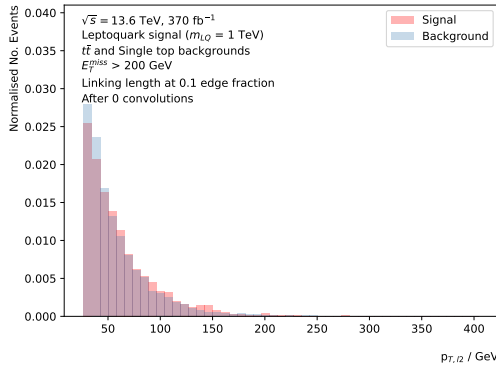
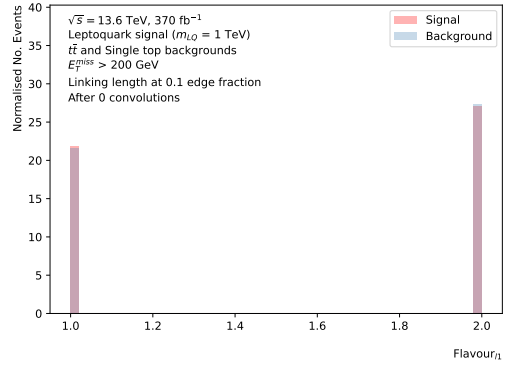
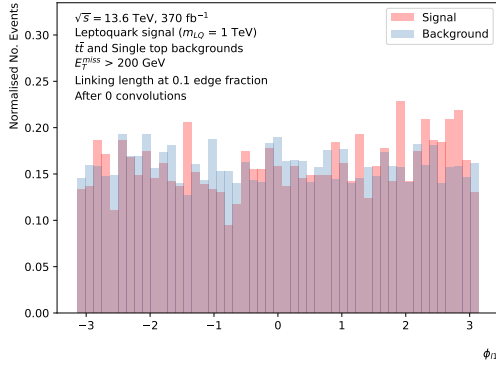
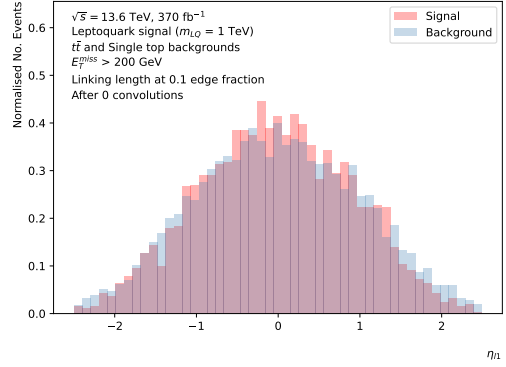
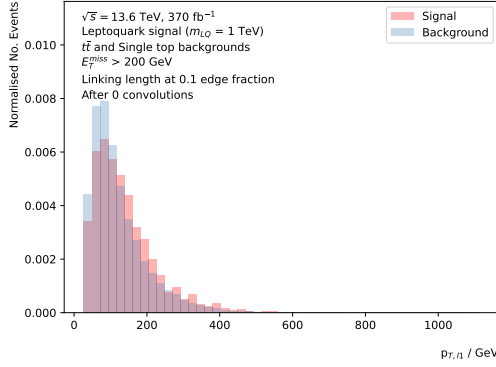
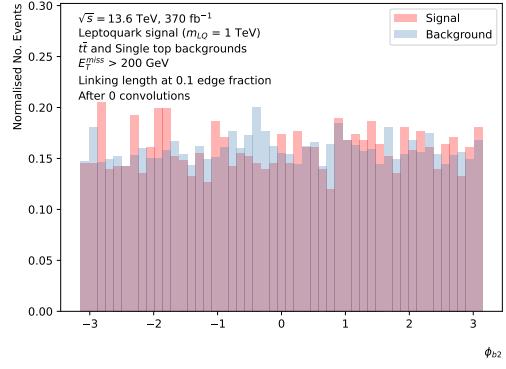
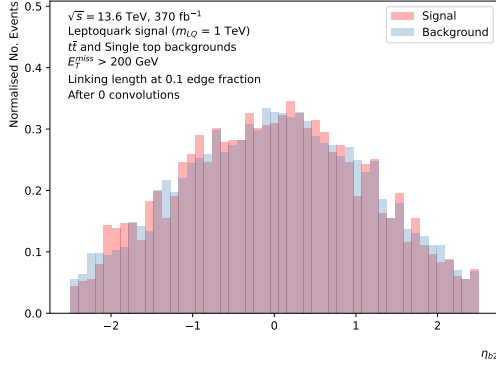


Figure 6.7: The Euclidean distances (a) and cosine distances (b) between events in a dataset consisting of LQ signal with $m_{LQ} = 1$ TeV and SM $t\bar{t}$ and single top backgrounds, calculated in a kinematic space constructed from 5 high-level kinematic variables: E_T^{miss} , $p_{T,bbll}$, $m_{T,l1}$, $m_{T,l2}$, and $\text{Min}\Delta\phi_{E_T^{\text{miss}},l1}$. The distances between signal and signal events are shown in blue, signal and background events in orange, and background and background events in green. All histograms are normalised to have unit area to allow for shape-only comparisons between the distributions. The KL divergence values between the three types of distance distributions are quoted to quantify their separation. The red dashed lines show several linking lengths and the corresponding fractions of edges formed.





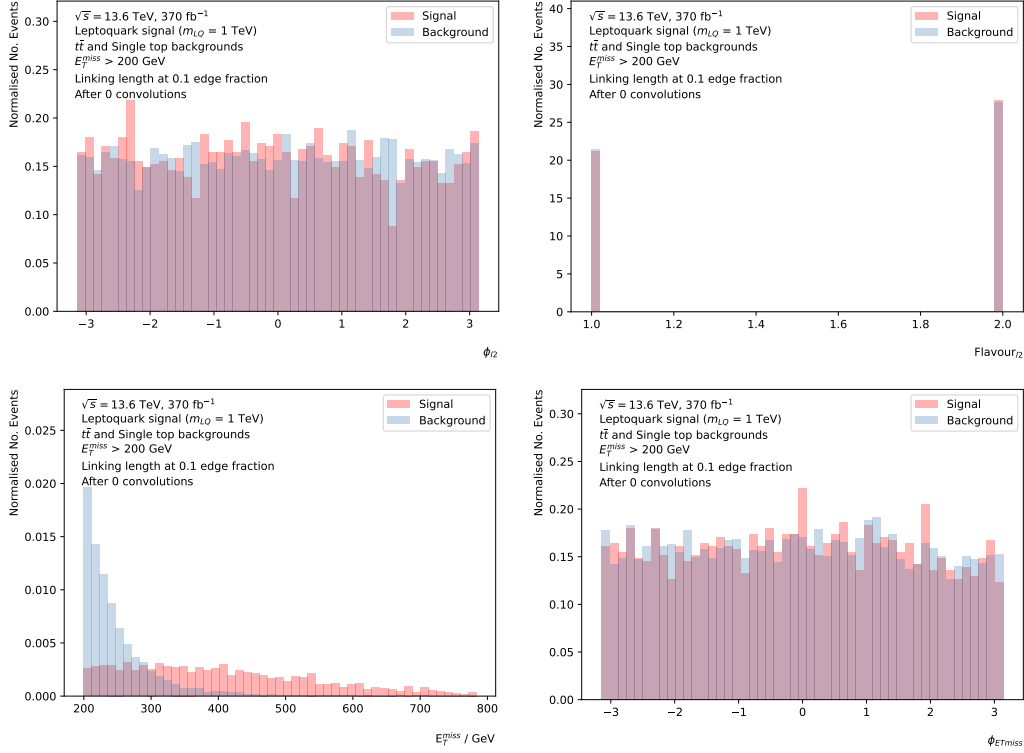


Figure 6.8: Distributions of the low-level kinematic variables, p_T , η and ϕ of the leading and sub-leading b -jets, the leading and sub-leading leptons, the flavour of the leptons, E_T^{miss} and its distribution in ϕ , between LQ signal with $m_{LQ} = 1$ TeV and SM $t\bar{t}$ and single top backgrounds. They form the 14-dimensional kinematic space where the distances between events can be calculated. All histograms are normalised to have unit area to allow for shape-only comparisons between the distributions.

6.3.3 Event Similarity in Latent Space

While the graph domain can be constructed using physical kinematic variables, it does not necessarily maximise the topological differences between signal and background events. Constructing a high-level kinematic space requires physics knowledge about the physics processes which could be challenging if the topological disparities exist only in high dimensionality. Whereas a low-level kinematic space, albeit general, often contains duplicate information and lacks physics motivations.

An alternative approach is to construct a domain in a *latent space* where events are embedded by a separate ML model, referred to as EmbeddingNet in this study. This technique falls under the field of *metric learning* [124], where a ML algorithm is trained to learn an optimal distance function in order to provide metrics based on data to facilitate classification. Intuitively speaking, the algo-

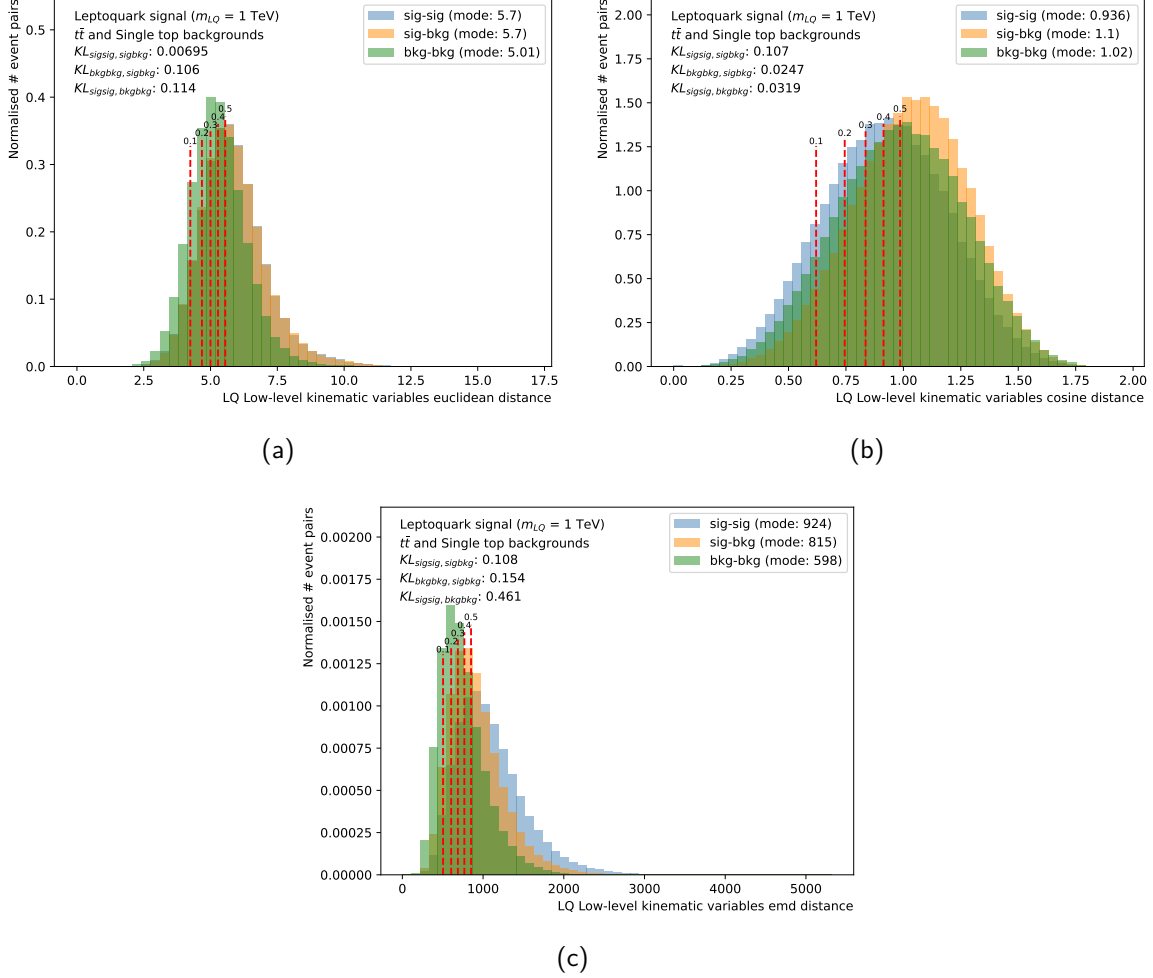


Figure 6.9: The Euclidean distances (a), cosine distances (b) and EMD distances (c) between events in a dataset consisting of LQ signal with $m_{LQ} = 1$ TeV and SM $t\bar{t}$ and single top backgrounds, calculated in a kinematic space constructed from 14 low-level kinematic variables: p_T , η and ϕ of the leading and sub-leading b -jets, the leading and sub-leading leptons, the flavour of the leptons, E_T^{miss} and its distribution in ϕ . The distances between signal and signal events are shown in blue, signal and combined background events in orange, and background and background events in green. All histograms are normalised to have unit area to allow for shape-only comparisons between the distributions. The KL divergence values between the three types of distance distributions are quoted to quantify their separation. The red dashed lines show several linking lengths and the corresponding fractions of edges formed.

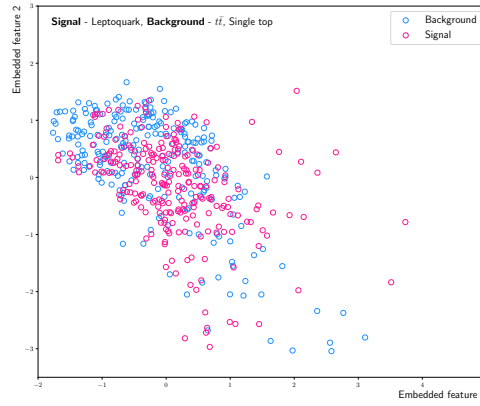
rithm performs clustering of datapoints by choosing metrics to form a latent space that naturally separates events of different classes. In the case of separating LQ signal and SM background events, the desirable latent space forms two local clusters for signal and background events that are far apart from each other. For the clustering of two classes, the contrastive hinge loss is used

$$\mathcal{L}(i, j, y) = y \cdot d_{i,j}^2 + \lambda(1 - y) \cdot \max[0, m - d_{i,j}]^2, \quad \text{where } y = \begin{cases} 1 & \text{if } C(i) = C(j) \\ 0 & \text{otherwise} \end{cases} \quad (6.14)$$

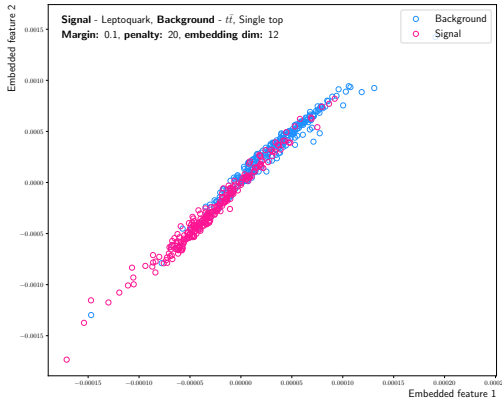
where y is the target label for each pair of events, $d_{i,j}$ is the Euclidean distance between events i and j , and $C(i), C(j)$ are their class labels. The parameter λ is a penalty term that penalises the closeness between different clusters. The parameter m is the *margin*, which is the minimal distance between two events that the loss function will tolerate. The first term of the loss function minimises the Euclidean distances between event pairs of the same class, and the second term maximises the Euclidean distances between event pairs of different classes with $d_{i,j} < m$. During training, the algorithm iteratively clusters same-class event pairs and pushes different-class events pairs apart. The outputs of the model are M values per-event, where M is the dimension of the latent space. Within it, events are connected using the same approach as in the kinematic space, where a pair of events i and j are connected if $d_{i,j} < LL$. In this case, the graph topologies are such that two distinct clusters are formed by events in the signal and background classes. This generalises the graph features for different types of signal models, removing the need for designing metrics using physics knowledge to connect nodes. Moreover, the training of this algorithm utilises low-level kinematic variables as input features, which requires no additional physics knowledge.

An EmbeddingNet model is trained using sampled LQ signal events with $m_{LQ} = 1$ TeV and SM $t\bar{t}$ and single top background events to construct a 12-dimensional latent space. The training details and the optimisations of the model parameters are detailed in Appendix F. To illustrate the clustering effect of the algorithm, distributions of 2 randomly selected features of signal and background events in a 12-dimensional latent space are shown before training, after 5 epochs and after 30 epochs of training in Figure 6.10. The distances between events in the resulting latent space demonstrate

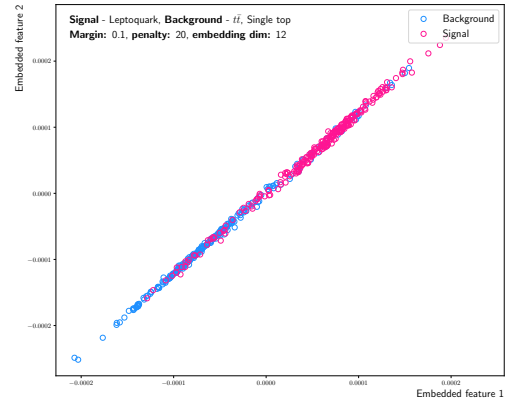
significantly improved disparity in graph topologies between the signal and background classes, seen in Figure 6.11. However, this approach leads to higher density of edges within the graph, which increases the chance of saturating the memory usage of the machine during the GNN training. This can be managed by adjusting the choice of linking length such that the edge fraction is kept to a low percentage. Other solutions such as neighbour sampling to tackle the challenge in the scalability of the graph representation are discussed in Section 6.4.



(a)



(b)



(c)

Figure 6.10: Distributions of 2 randomly selected features of LQ signal events with $m_{LQ} = 1$ TeV (pink) and SM $t\bar{t}$ and single top background events (blue) in a 12-dimensional latent space. The distributions are shown before training (a), after 5 epochs (b) and after 30 epochs of training (c). A margin of 0.1 and a penalty of 20 relating to the contrastive hinge loss function were used as training parameters. The clustering effect between same-class events and the separation between the signal and background classes are clearly evident.

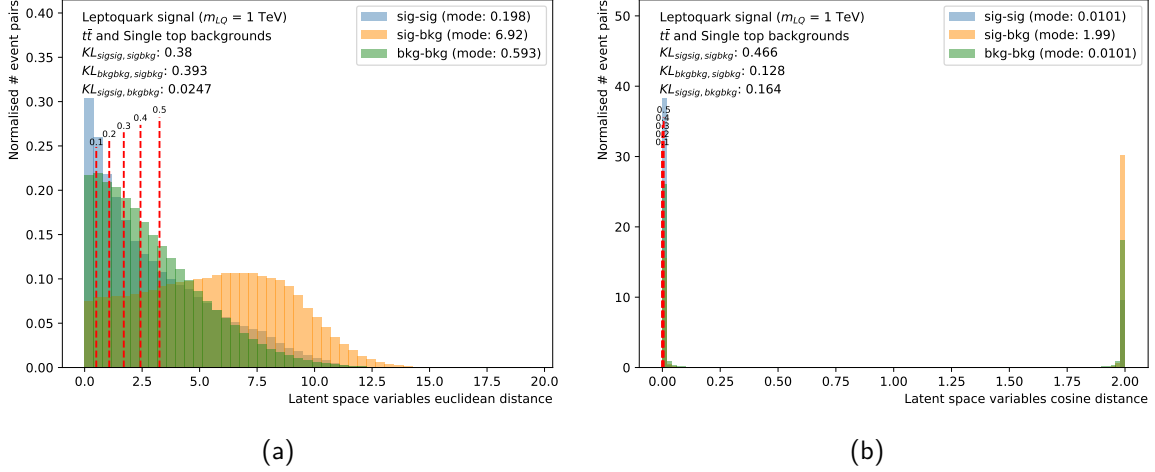


Figure 6.11: The Euclidean distances (a) and cosine distances (b) between events in a dataset consisting of LQ signal with $m_{LQ} = 1$ TeV and SM $t\bar{t}$ and single top backgrounds, calculated in a 12-dimensional latent space constructed using the outputs of an ML clustering algorithm EmbeddingNet. The distances between signal and signal events are shown in blue, signal and background events in orange, and background and background events in green. All histograms are normalised to have unit area to allow for shape-only comparisons between the distributions. The KL divergence values between the three types of distance distributions are quoted to quantify their separation. The red dashed lines show several linking lengths and the corresponding fractions of edges formed.

6.3.4 Kinematic Convolution

A set of per-node kinematic variables $\mathbf{X}$ are used as input features for the GNN training. In each hidden layer of the neural network, they are convoluted with those of their nearest neighbours, referred to as message passing. The results of the convolution depend on the graph connectivity, which is determined by the domain within which the graph is constructed, the distance metrics used, the target edge fraction used to define the linking length, as well as the graph convolution function, detailed in Section 6.2.4. The results are convoluted kinematic variables, which are then combined with trainable parameters before inputted to a non-linear activation function, expressed in Equation 6.7. To visualise the effect of the convolution, a given convoluted kinematic variable is shown before being combined with the trainable parameters. For example, the original, once convoluted and twice convoluted p_T distributions of the leading b -jet of the LQ signal with $m_{LQ} = 1$ TeV and the SM background events are shown in Figure 6.12. Here, the graph is constructed in the low-level kinematic domain using the EMD distance with an edge fraction of 0.2 and the convolution function for Graph Convolution Network in Equation 6.9. Other convoluted kinematic

variables using graphs constructed in the high-level kinematic domain and the latent space, using various distance metrics can be found in the Appendix G.1.

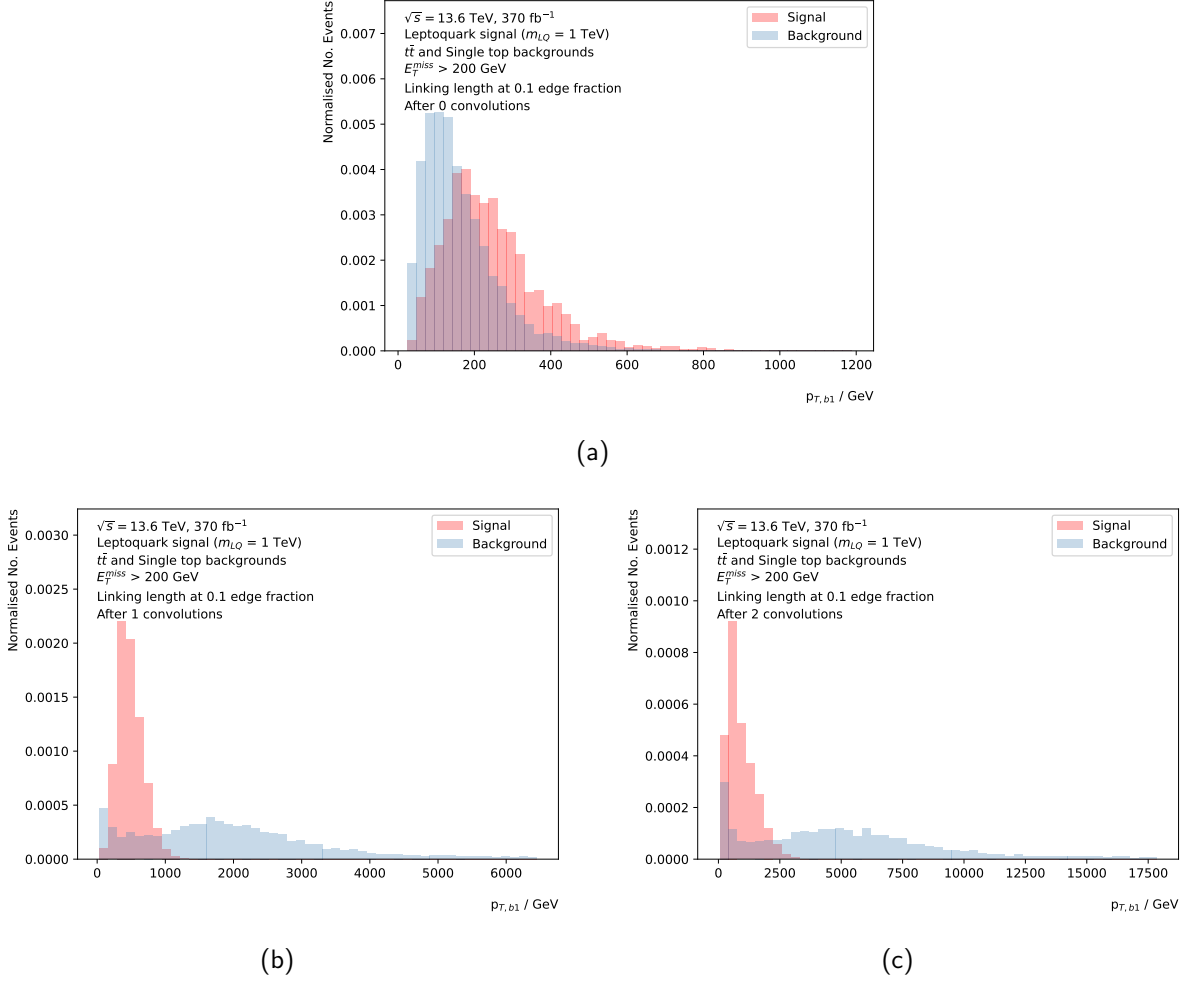


Figure 6.12: Distributions of the convoluted p_T of the leading b -jet in the Leptoquark signal with $m_{LQ} = 1 \text{ TeV}$ and the SM $t\bar{t}$ and single top backgrounds. The distributions with no convolution (a), after 1 convolution (b) and after 2 convolutions (c) are shown. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in the low-level kinematic domain using the Earth Mover's Distance metric with a linking length corresponding to a target edge fraction of 0.1. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

6.4 Model Scalability

In many particle physics analyses, large simulated samples of various physics processes are used to train ML models. This poses a scalability challenge for these models, which are often constrained by the memory required for computational storage during their training. The size of a graph repre-

presentation of an entire dataset directly scales with its size as the adjacency matrix scales as $\mathcal{O}(N^2)$. This leads to a large number of node and edge features to be stored computationally. In message passing, the number of neighbours included in the convolution also increases exponentially as the number of hidden layers increases, leading to a large number of computations. One could reduce the size of the graph representation by limiting the linking length to be small such that fewer edges between nodes are formed. There are also various computational solutions to the memory limitations encountered by large GNN models, ones adopted in this study include

- **Sparse Tensors:** storing only the indices of the non-zero values in the adjacency matrix in a sparse format.
- **Neighbour sampling:** sampling a subset of the nearest neighbours and their subsequent neighbours to a fixed depth for each node.
- **Mini-batch training:** loading the graph representation in batches during training.

The details of each technique are briefly discussed here.

Sparse Tensor matrix encoding is used to reduce the memory required to store the relevant information from the adjacency matrix. Instead of storing $(N \times N)$ entries of the entire matrix, only the row and column indices of the non-zero entries are stored and loaded. This reduces the size of the adjacency matrix to scale with $\mathcal{O}(E)$, which is managed by the choice of the linking length.

Neighbour sampling is a method initially proposed in the GraphSAGE [125] framework for inductive learning. A maximum number of neighbours are randomly sampled in each neighbourhood layer starting from a target node to a specified depth within its neighbourhood. It relies on the model learning an embedding function that can generalise node features as well as topological features within neighbourhoods to unseen nodes. By randomly sampling a subset of the neighbours for each target node, it also acts as a form of regularisation similar to dropout, which prevents the model overfitting features that are specific to the graph representation of the training dataset. Figure 6.13 demonstrates an example of neighbour sampling for a target node, with a sampling scheme of $[3, 3]$. In other words, a maximum of 3 nearest neighbours of the target node are sampled, then a maxi-

mum of 3 of their nearest neighbours are further sampled. Sampling deeper into the neighbourhood captures more global features the graph, whilst sampling a higher number of nodes in each layer captures more local features within the neighbourhood.

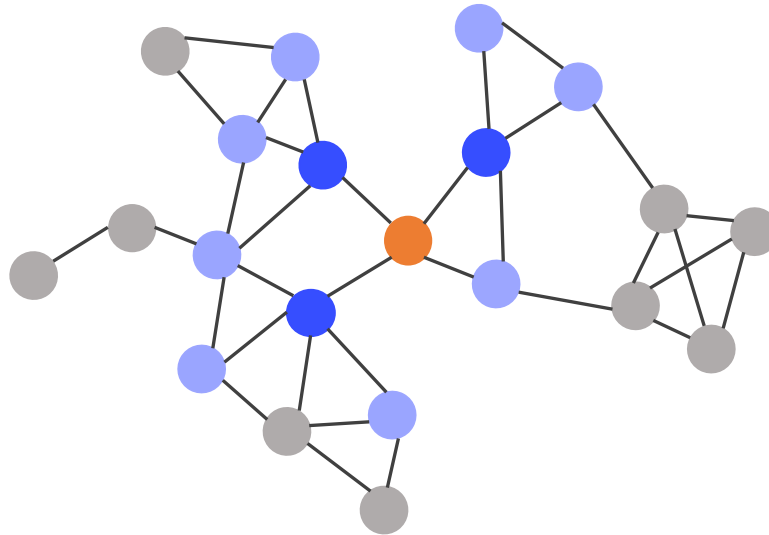


Figure 6.13: A schematic demonstrating neighbour sampling. The target node is in orange, and a maximum of 3 nearest neighbours are sampled to the a neighbourhood depth of 2. The nodes sampled in the first layer are shown in deep blue, and the ones sampled in the second layer are in light blue. The nodes that are not sampled are in grey.

Mini-batch training allows the model parameters to be updated over small sub-graphs consisting of a subset of target nodes and their sampled neighbours. The message passing is only computed over each sub-graph, and the loss calculated for only the target nodes within it to update the trainable parameters. The message passing and parameter update are iterated over all mini-batches until all the target nodes and their sub-graphs within the graph representation have been included. This allows all features of the graph to contribute to the training without having to store the entire graph in memory.

6.5 Model Implementation and Training

An array of GNN models are trained to evaluate the impacts of the following parameters: the input node features, the graph domain within which the graph representation is generated, the distance metric used to define the linking length, and the convolution model which differs in the convolution

function used in the message passing. Table 6.1 summarises the setup for each set of GNNs trained to compare the GCN and GraphConv convolution models.

Kinematic features	Graph domain	Distance metric	Convolution model
High-level	High-level kinematic space	Euclidean, cosine	GCN, GraphConv
Low-level	Low-level kinematic space	Euclidean, cosine, EMD	GCN, GraphConv
Low-level	Latent space	Euclidean, cosine	GCN, GraphConv

Table 6.1: Table summarising the setup for each set of GNNs comparing the GCN and GraphConv convolution models. The varying elements include the per-node kinematic features used as input features, the domain and distance metric used to construct the graph, and the convolution model used in the GNN training.

The signal class consists of roughly 2500 simulated events of the LQ signal sample with $m_{LQ} = 1$ TeV, and the background class consists of roughly 8000 events from the combined SM $t\bar{t}$ and single top production samples, after selection criteria are applied. The classification of these two classes is performed by using a binary cross entropy loss function

$$L(y, \hat{y}) = -\frac{1}{\sum_i^N w_i^{MC}} \sum_i^N \left\{ w_i^{MC} * w_i^{class} [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \right\} \quad (6.15)$$

where y is the target label, which for a binary classification is 0 and 1 for the two classes, $\hat{y}$ is the model prediction for each event, and N is the number of samples in each mini-batch over which the loss is averaged in each training iteration. Each event's contribution to the total loss is weighted by a class weight w_i^{class} , and its MC simulation weight w_i^{MC} . The class weight is $\frac{N_{background}}{N_{signal}}$ for signal events, and its inverse for background events. It accounts for the statistical imbalance between the signal and background classes by assigning more weight to the contribution from signal events in the loss calculation. The MC simulation weight, as introduced previously, represents an event's effective yield taking into account the theory predicted cross-section of the underlying physics process, detector luminosity and efficiencies. The assignment of these weights ensures that the trained model does not bias towards the class with larger available statistics, and is robust when applied to recorded data. The k -fold splitting procedure with $k = 3$ is used for the training of each model in order to prevent overtraining while making full use of the available training statistics. The training and validation losses as functions of the number of epochs can be found in Appendix G.2 for the all

models.

The architectures of the baseline DNN and various GNN models are summarised in Table 6.2. Two baseline DNN models are trained, one using high-level kinematic variables as input features, and the other using low-level kinematic variables. Dropout regularisation is applied during the DNN training to mitigate the effect of overtraining. For the various GNN models, a combination of GNN layers involving message passing between nodes and fully connected, feedforward multilayer perceptron (MLP) layers is used. There is an exception for the GraphConv models that use graphs defined using cosine distance, as it was found that using only GNN layers yields more stable model convergence. The edge fraction that corresponds to the chosen linking length is chosen to be either 0.1 or 0.2 to accommodate the memory limitation of the computing infrastructure used for training. Neighbour sampling is adopted to further reduce the density of the graph with minimal loss of the features in the graph representation by sampling a large number of nodes to a relatively large depth.

6.5.1 Transductive Learning

In order to capture both local and global features within the graph representation of the full dataset, *transductive learning* [126], [127] is used where the entire graph, consisting of both the training and validation datasets, is visible to the GNN model during training. This differs from *inductive training*, which is more common in the training of supervised learning algorithms, where the dataset is split into training and validation sets, and only the training set is used to train the model. In the case of GNNs, splitting the dataset effectively samples the graph representation, creating subgraphs that may not be representative of a real dataset. Transductive learning allows the model to learn features from the entire graph, leveraging the fact that nodes forming similar patterns often belong to the same class. Reference [127] also shows better generalisation that is consistent with the theoretical upper bound compared to inductive learning. During the GNN training, the model learns features through message passing between neighbouring nodes over the entire graph. At the back propagation step, predictions are made only for nodes in the validation dataset which then contribute to the loss calculation and the update of the trainable parameters within the model. However, the nodes in the validation dataset still have dependency on the entire graph, as they may be connected

Convolution model	Distance metric	Graph domain	GNN layers	MLP layers	Edge fraction	Neighbours sampled [nodes, layers]	Dropout
High-level kinematic input variables							
DNN				[12, 12, 12, 12]			0.05
GCN	Euclidean	High-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.2	[60, 6]	0.0
GCN	Cosine	High-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.2	[60, 6]	0.0
GraphConv	Euclidean	High-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.2	[60, 6]	0.0
GraphConv	Cosine	High-level kinematic space	[32, 32, 32, 32]		0.2	[20, 2]	0.0
GCN	EMD	Low-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.2	[60, 6]	0.0
GraphConv	EMD	Low-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.2	[60, 6]	0.0
Low-level kinematic input variables							
DNN				[12, 12, 12, 12]			0.1
GCN	Euclidean	Low-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GCN	Cosine	Low-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GCN	EMD	Low-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GraphConv	Euclidean	Low-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GraphConv	Cosine	Low-level kinematic space	[12, 12]		0.1	[20, 2]	0.0
GraphConv	EMD	Low-level kinematic space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GCN	Euclidean	Latent space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GCN	Cosine	Latent space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GraphConv	Euclidean	Latent space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0
GraphConv	Cosine	Latent space	[32, 32, 32, 32]	[12, 12]	0.1	[60, 6]	0.0

Table 6.2: Table summarising the neural network architecture used in the training for each variation of the baseline DNN and GNN models.

to nodes in the training dataset, hence are representative of the model's performance over the entire dataset.

6.6 Model Performance

Model performance is evaluated using the validation dataset, which is not used to perform gradient descent, but is still present to preserve the full graph topology during training. The performance is characterised by the following metrics:

- The Area Under the ROC curve (AUC): a value obtained by integrating over all true positive rates of the model outputs, giving an overview of the model performance.
- The Z significance: a value which quantifies the sensitivity of a targeted signal over background in signal regions defined to be above certain thresholds in the model output scores. It is calculated using Equation D.1. It assumes a 20% uncertainty on the expected number of background events, which roughly corresponds to the size of the systematic uncertainties in Reference [115].

6.6.1 Area Under the ROC Curve (AUC)

The AUCs of a baseline DNN, and GCN and GraphConv models, trained using high-level and low-level kinematic variables as input features are summarised in Table 6.3. For GNN models trained using high-level kinematic input features, the graphs are constructed in the high-level kinematic space using either the Euclidean or cosine distance metrics, or in the low-level kinematic space using EMD. For GNN models trained using low-level kinematic input features, the graphs are constructed either in the low-level kinematic space using all three distance metrics, or the latent space using the Euclidean or cosine distance.

For models trained with high-level kinematic input features, the baseline DNN and the GNN models achieve similar levels of performance. The GraphConv model with the EMD outperforms the baseline DNN, indicating the topological features in the associated graph representation offer additional

discriminating power. This is correlated with the largest KL divergence provided by the EMD between sig-sig and bkg-bkg distances using in the low-level kinematic space shown in Figure 6.9(c), compared to other distance metrics. Intuitively speaking, the graph topologies are such that SM background events are more clustered whilst the signal events are more sparsely distributed in the graph domain.

For models trained with low-level kinematic input features, their overall performances are worse than ones trained with high-level kinematic variables, which is unsurprising as the input features are less discriminating. For models using graphs constructed in the low-level kinematic space and the Euclidean and cosine distances show no AUC gain over the baseline DNN. However, models using the EMD yield higher AUC values compared to the baseline DNN, for both the GCN and GraphConv models. Again, this points to the additional discriminating power in graph topologies provided by the EMD. For GNN models using graphs constructed in the latent space, their AUCs are higher than the ones using graphs constructed in the low-level kinematic space in general. However no gain in performance is observed compared to the baseline DNN.

6.6.2 Z Significance

In particle physics, the Z significance is often calculated to quantify the number of standard deviations that the observed data deviates from a hypothesis in frequentist hypothesis testing. In contrast to the AUC values, they show the expected sensitivity to a particular physics signal that can be achieved by defining a signal region above certain thresholds in the machine learning model output scores and performing frequentist testing.

The Z significances are calculated for the LQ signal with $m_{LQ} = 1$ TeV and the combined $t\bar{t}$ and single top background events, scanning across the output scores of the trained models. The baseline DNN and GNN models trained with high-level kinematic input features are shown together in Figure 6.14. The comparison is made between the different distance metrics in the high-level and low-level kinematic space, while fixing the input features. Whilst the AUC values do not show significant performance gain from the GNNs trained using high-level kinematic inputs compared to

Convolution model	Distance metric	Graph domain	AUC (validation)
High-level kinematic input variables			
DNN			0.956
GCN	Euclidean	High-level kinematic space	0.951
GCN	Cosine	High-level kinematic space	0.954
GraphConv	Euclidean	High-level kinematic space	0.943
GraphConv	Cosine	High-level kinematic space	0.952
GCN	EMD	Low-level kinematic space	0.954
GraphConv	EMD	Low-level kinematic space	0.972
Low-level kinematic input variables			
DNN			0.919
GCN	Euclidean	Low-level kinematic space	0.826
GCN	Cosine	Low-level kinematic space	0.852
GCN	EMD	Low-level kinematic space	0.901
GraphConv	Euclidean	Low-level kinematic space	0.812
GraphConv	Cosine	Low-level kinematic space	0.804
GraphConv	EMD	Low-level kinematic space	0.951
GCN	Euclidean	Latent space	0.892
GCN	Cosine	Latent space	0.901
GraphConv	Euclidean	Latent space	0.892
GraphConv	Cosine	Latent space	0.873

Table 6.3: Table summarising the Area Under the ROC curves (AUC) values evaluated on the validation dataset, for the baseline DNNs and various GNN models, trained using the high-level or low-level kinematic variables as input features.

the baseline DNN, the GraphConv model using the EMD yields a higher Z significance over the DNN consistently in the high score region. The GCN model using the Euclidean distance also outperforms the baseline DNN in the very high score region, attributed to a larger excess of signal events in the very last bin over background.

The baseline DNN and GNN models trained with low-level kinematic input features are shown together in Figure 6.15, where their graphs are constructed either in the low-level kinematic space or the latent space, using varying distance metrics. Here, only the GraphConv model using a graph constructed in the low-level kinematic space using the EMD surpasses the baseline DNN in the high score region. GNN models using graphs constructed in the latent space show no performance gain, despite using a dedicated algorithm to engineer a graph domain where the signal and background events form two separated clusters. This suggests that the clustering of same class events in the graph representation does not necessarily aid the signal background discrimination in the GNN training. Whereas the EMD, which yields more clustered background events and more sparse signal events in the graph domain, is the more effective at signal background discrimination.

Lastly, the baseline DNN and GNN models trained with low-level kinematic input features, using graphs constructed in the low-level kinematic space are shown separately for the Euclidean and the cosine distance metrics in Figure 6.16. Here, no GNN models exceed the baseline DNN. This is likely due to the less discriminating distances calculated in the low-level kinematic space, which results in less separation between signal and background in the convoluted kinematics compared to the original distributions.

6.7 Conclusions

The feasibility and performance of a novel approach to search for new physics at the LHC using dataset-wide Graph Neural Networks are explored. The method aims to exploit the topological differences between BSM signal and SM background events within a graph representation of the whole dataset. Several factors are experimented, the graph domain and distance metrics used to derive the

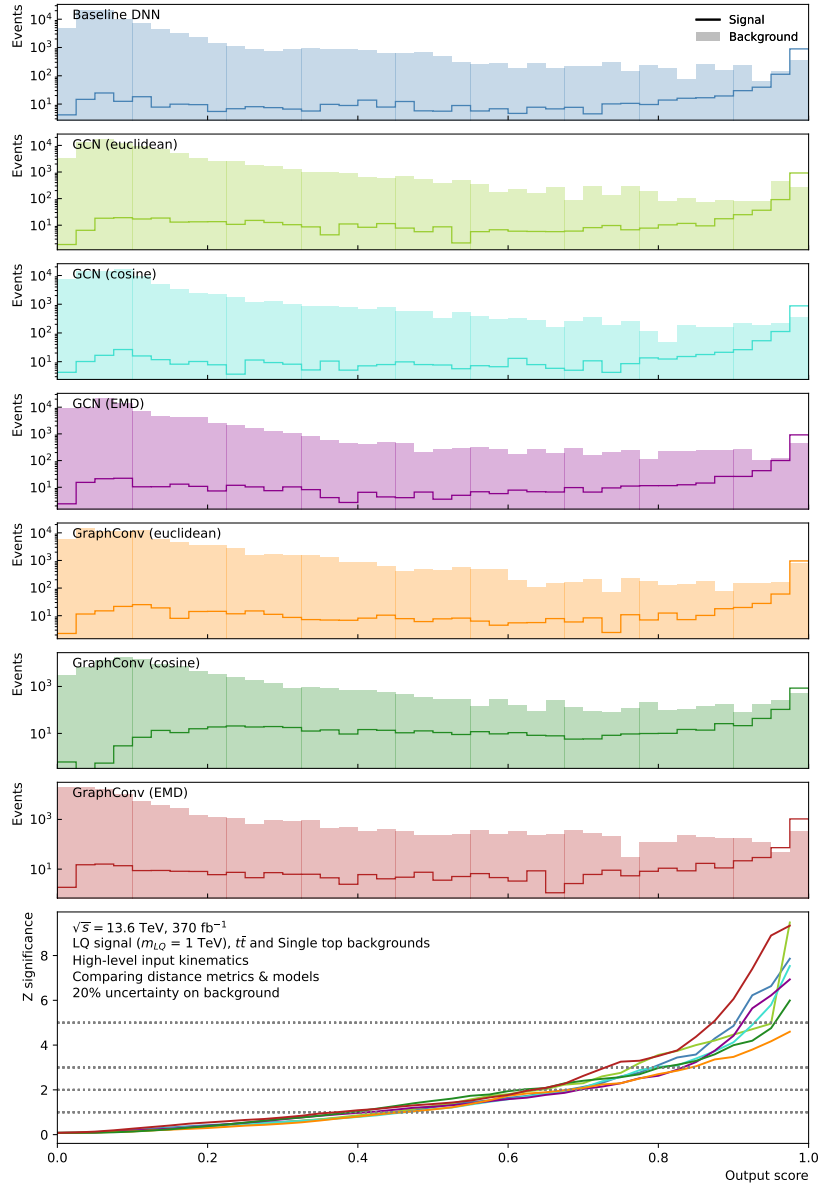


Figure 6.14: The output score distributions and the Z significance of the LQ signal events with $m_{LQ} = 1$ TeV and the combined SM $t\bar{t}$ and single top background events above the score value at each bin. Events are weighted by their MC event weights to represent the expected yields with the theory predicted cross-sections accounted for. All models are trained with high-level kinematic variables. The comparison between the GNN models are made between ones trained with graphs defined in the high-level kinematic space using Euclidean or cosine distance, and ones trained with graphs defined in the low-level kinematic space using EMD. The GCN model using the Euclidean distance and the GraphConv model using the EMD yield a higher Z significance than the baseline DNN at the high score region. The values of $Z = 1, 2, 3$, and 5 which corresponds to the discovery sensitivity are indicated by dashed lines in the bottom panel.

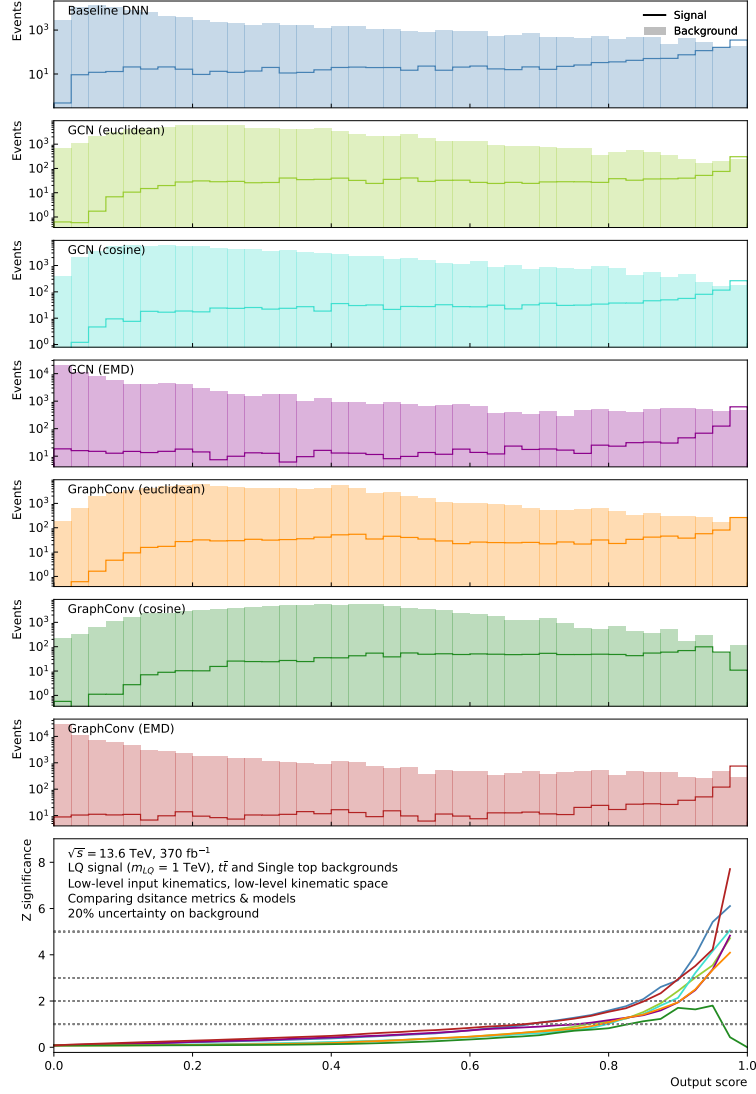


Figure 6.15: The output score distributions and the Z significance of the LQ signal events with $m_{LQ} = 1$ TeV and the combined SM $t\bar{t}$ and single top background events above the score value at each bin. Events are weighted by their MC event weights to represent the expected yields with the theory predicted cross-sections accounted for. All models are trained with low-level kinematic variables. The comparison between the GNN models are made between ones trained with graphs defined in the low-level kinematic space using Euclidean distance, cosine distance or the EMD. The GraphConv model using the EMD yields a higher Z significance than the baseline DNN at the high score region. The values of $Z = 1, 2, 3$, and 5 which corresponds to the discovery sensitivity are indicated by dashed lines in the bottom panel.

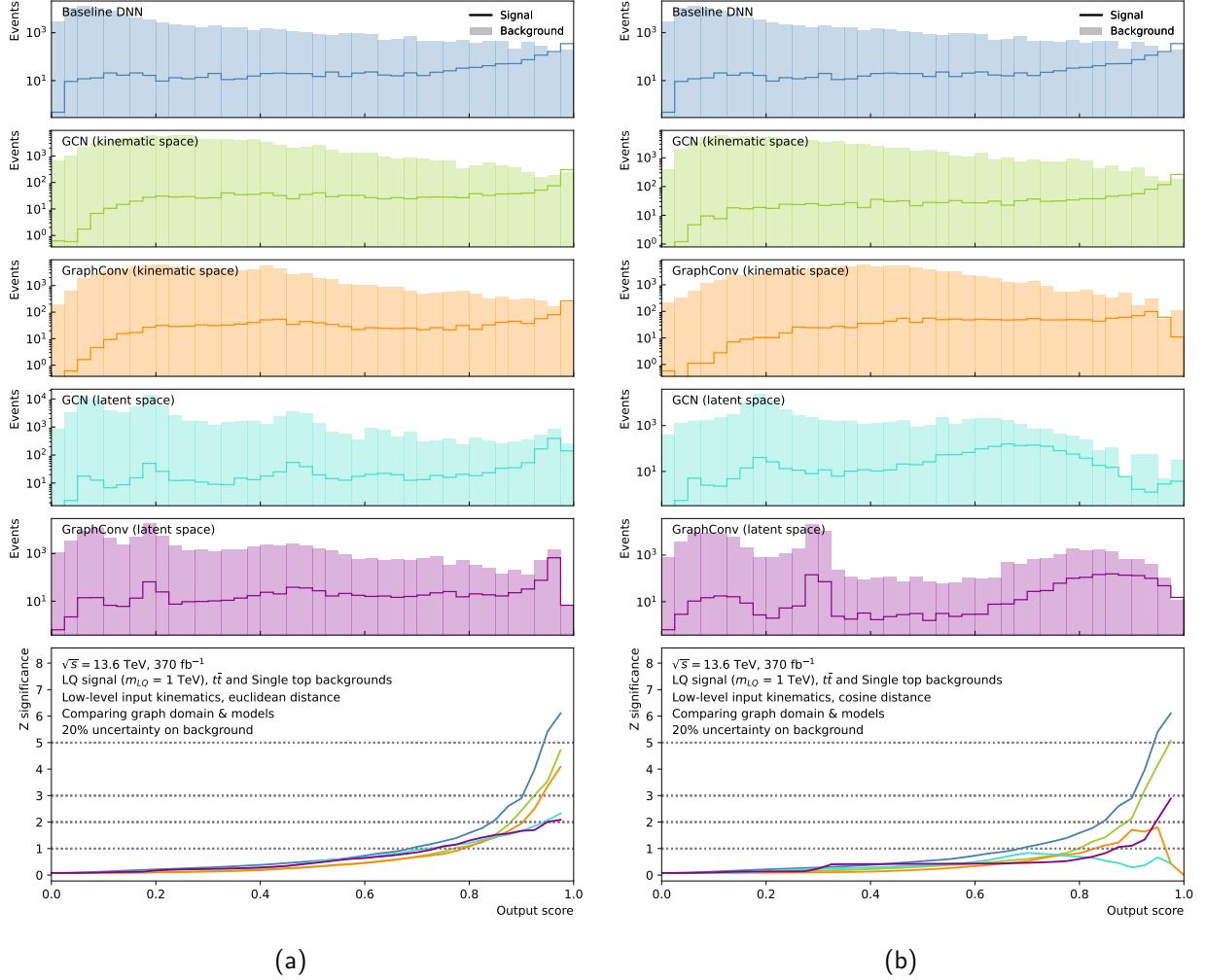


Figure 6.16: The output score distributions and the Z significance of the LQ signal events with $m_{LQ} = 1$ TeV and the combined SM $t\bar{t}$ and single top background events above the score value at each bin. Events are weighted by their MC event weights to represent the expected yields with the theory predicted cross-sections accounted for. All models are trained with low-level kinematic variables. The comparison between the GNN models are made between ones trained with graphs defined using the Euclidean distance (a) or the cosine distance (b) in the low-level kinematic space and the latent space. For both distance metrics, no GNN models outperform the baseline DNN. The values of $Z = 1, 2, 3$, and 5 which corresponds to the discovery sensitivity are indicated by dashed lines in the bottom panel.

graph representation, the kinematic variables used to train the models, and the convolution function the GNN model uses to perform message passing.

Three different graph domains are explored. One is constructed using high-level kinematic variables, which are highly discriminating but require physics knowledge. Another is constructed using low-level kinematic variables, which are more general but are less discriminating. Alternatively a latent space is formed using outputs from a dedicated algorithm that clusters same-class events. Unsurprisingly, the graphs constructed in the high-level kinematic space yield more performant GNNs, but are less general as specific physics knowledge about the targeted signal is required to select these kinematic variables. The low-level kinematic space trades off the model performance with the potential to generalise over a wider range of final states and signals. To compensate for this loss of performance, an alternative approach using a clustering neural network, the `EmbeddingNet`, to derive a high-dimensional latent space from the low-level kinematic variables is explored. The graph representations derived in this latent space using either the Euclidean or cosine distance yield more performant GNNs in terms of their AUCs, but provide no sensitivity gain in terms of the Z significance compared to other models trained with the low-level kinematic inputs.

For distance metrics, the study experimented with using Euclidean and cosine distances to derive the linking length for graph construction within the high-level and low-level kinematic space, and the latent space. The cosine distance in general provides better discrimination between signal and background, as the LQ signal events are more angularly separated due to the large number of invisible particles in the final states causing the kinematic variables to be more varied. A more physically motivated metric, the Earth Mover's Distance (EMD), is also explored within the low-level kinematic space. It yields larger distances between signal events and small distances between background events, which further accentuates the large variance in the LQ signal kinematics. As a result, the GNN models that use the corresponding graph representations, in both the low-level and high-level kinematic spaces, have the best performance both in terms of AUC and Z significance. They also outperform the corresponding baseline DNNs, proving that the topological features present in the dataset-wide graph representation may contribute to the separation of BSM signal and SM

background events.

As machine learning becomes an increasingly common tool in the searches for new physics, it is crucial to expand the horizon of its development and application. Whilst DNNs are often sufficiently performant and exceed the signal sensitivity from using a traditional cut and count strategy, they can be further improved or superseded by more innovative approaches. This study presents one such approach, where the architecture is inspired by the latest advancement in graph-based machine learning models, and the concept remains strongly motivated by physics. The aim is to improve the chances of new physics discovery, while preserving the model's ability to generalise to a wide range of signatures and its physics interpretability. This is achieved by the combination of using low-level kinematic variables to construct the graph representation and as input features for the GNN training, and using a physically motivated distance metric, the EMD. The performance gain over the baseline approach is promising, which prompts future studies regarding the validation and application in a physics analysis. This study also presents the first attempt of using a dataset-wide ML approach for an LHC analysis.

6.8 Future Studies

Aiming to further improve the performance of this approach, several studies are either currently underway or planned for the future. One explores the use of the *attention mechanism* within the GNN framework, namely the Graph Attention Network (GAT) [128]. The mechanism introduces the idea of importance as a trainable parameter, and weights each input accordingly during the training. It is commonly applied to models that deal with sequences as inputs, such as natural language processing and time series analysis. It also has the benefit of being able to deal with sequences of variable sizes. In particular, the *self-attention* [75] shows impressive ability to capture complex relationships within a sequence, where each element considers all other elements (including itself) to determine their importance when the representation of the element is computed. The importance is quantified by an attention score. Inspired by this concept, self-attention mechanism can be incorporated into a GNN architecture via an *attention layer*. When the features of a given

node are convoluted with its nearest neighbours' during message passing, the importance of the updates received from each neighbour can be learned by assigning an *attention coefficient* to each. The coefficient α_{ij} for a neighbour j of a target node i in the l^{th} hidden layer can be calculated as

$$\alpha_{ij} = \frac{\exp\left(\phi\left(\tilde{\mathbf{a}}^T[\mathbf{W}^{(l)}\vec{x}_i^{(l-1)} \parallel \mathbf{W}^{(l)}\vec{x}_j^{(l-1)}]\right)\right)}{\sum_{\nu \in \text{Neighbourhood}(i)} \exp\left(\phi\left(\tilde{\mathbf{a}}^T[\mathbf{W}^{(l)}\vec{x}_i^{(l-1)} \parallel \mathbf{W}^{(l)}\vec{x}_\nu^{(l-1)}]\right)\right)} \quad (6.16)$$

where $\vec{x}^{(l-1)}$ are node features from the previous hidden layer, and $\mathbf{W}^{(l)}$ are the learnable training weights in the l^{th} hidden layer. The attention mechanism is parametrised by an attention weight vector $\tilde{\mathbf{a}}$ that is combined with the concatenated node features $[\mathbf{W}^{(l)}\vec{x}_i^{(l-1)} \parallel \mathbf{W}^{(l)}\vec{x}_j^{(l-1)}]$. Here the attention weight becomes a trainable parameter that is updated for each iteration of training as it is a function of learnable training weights $\mathbf{W}$. The function ϕ is a non-linear activation function, in the case of GAT implemented in PyTorch Geometric [110] library, a LeakyReLU [129] function is used. The convolution of the features of nodes neighbouring the target node i is then performed

$$\vec{x}_i^{(l)} = \sigma\left(\sum_{j \in \text{Neighbourhood}(i)} \alpha_{ij} \mathbf{W}^{(l)}\vec{x}_j^{(l-1)}\right) \quad (6.17)$$

where σ is a non-linear activation function of the hidden layer. The message passing for this operation is illustrated in Figure 6.17.

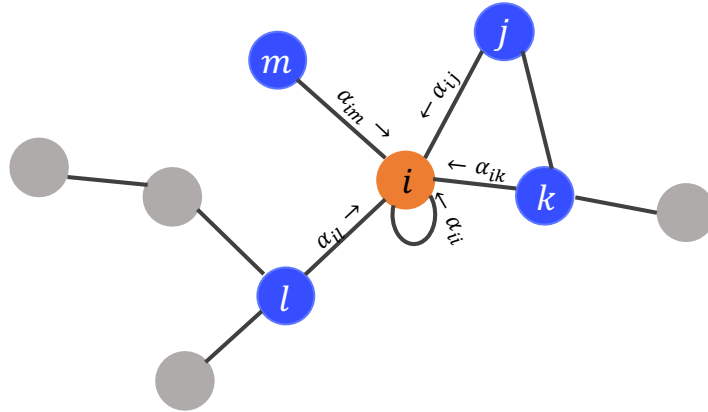


Figure 6.17: A schematic of the message passing of an example graph in the GAT network, where the features of a target node i in orange receive updates from all its nearest neighbours (in blue). The attention coefficients $\vec{\alpha}$ quantifying the importance of the update received from each nearest neighbour are updated during each iteration of the training.

This approach is highly efficient in convolution computation, as the calculation of the attention coefficients in Equation 6.16 can be parallelised across all edges, and the feature convolution in

Equation 6.17 can be parallelised across all nodes. The additional awareness of the relative importance between nearest neighbours allows the model to place higher importance on neighbours that contain more discriminating power.

The next step of the project involves validating the GNN models for robustness when applied to data. In a graph representation, events are no longer treated as independent as for the case in a DNN, therefore models are sensitive to changes in the graph topologies. The careful application of MC event weights, outlined in detail in Section 6.5, ensures that the graphs constructed using simulated events during training well represent the topologies expected in data. The model response will be tested on open source data events in control regions defined adjacent to the signal region. Closure will be checked between data and simulated MC samples in these control regions. Furthermore, the robustness of the model to background processes will be studied. The GNN model outputs of background events evaluated on a background-only dataset, and on a dataset consisting of both signal and background events will be validated. The desired behaviour is for the model outputs for background events to remain unchanged with the presence of signal events in the graph representation.

It is also crucial that a model trained on one LQ mass has reasonable performance on other masses in a general search for LQs. The trained GNN models will be applied to other masses, and the expected sensitivity to them will be quantified using the Z significance similar to the approach in Section 6.6.2. A potential improvement to the GNN approach would also be to add a functionality to parametrise the models on the mass of the targeted BSM particle. The main motivation for this is to allow for a smoother interpolation between the masses, discussed in more details in Section 5.2 in the previous chapter.

7 | Conclusions

I love Shrek ²¹.

— Sebastian Rutherford Colmenares

This thesis presents the first search for triple Higgs boson (HHH) production conducted at the Large Hadron Collider, using 126 fb^{-1} of proton-proton collision data at $\sqrt{s} = 13 \text{ TeV}$ recorded at the ATLAS experiment during Run-2 data-taking. To provide theoretical and experimental contexts, Chapter 1 outlines the SM, focusing on the Higgs mechanism and the Higgs self-coupling, as well as the feasibility to address open questions within the extended Higgs sectors. Chapter 2 lays out the setup of the principle experiment, the ATLAS experiment at the LHC, which is used to conduct the searches presented in this thesis. With a strong focus on the trigger system, the detector apparatus and the reconstruction of the relevant physics objects are described. Continuing with this focus, the study of a machine learning algorithm dedicated to rejecting a common background arising from $g \rightarrow b\bar{b}$ splitting in the b -jet triggers is presented in Chapter 3. A Deep Neural Network (DNN) is developed to dedicatedly separate single- b -jets from bb -jets. When used in conjunction with the baseline b -tagging algorithm in the Run-3 b -jet triggers GN1, a higher bb -jet rejection rate is achieved. A procedure is developed to scan through the working points of both taggers to find the optimal combinations for good levels of rejection across all background jet flavours. The algorithm is able to reduce the readout rates and improve the multijet background rejection of a trigger used in the non-resonant $HH \rightarrow 4b$ analysis, pointing to the potential increase in signal efficiency in the future.

The search for the triple Higgs boson production in the $b\bar{b}b\bar{b}b\bar{b}$ final state takes advantage of the dominant Higgs boson decay mode to b -quarks, and targets both the resonant and non-resonant channels, in the gluon-gluon fusion (ggF) production mode. The non-resonant channel includes the SM, and BSM HHH production when one of the scalar particles is off-shell, whereas the resonant channel only includes the BSM HHH production when both scalar particles are on-shell. The event kinematic features of the two channels differ significantly, they are therefore interpreted using separately trained neural networks and background estimates, but following the same search

¹In Latin American Spanish.

strategy. The analysis places constraints on modifications to the trilinear and quartic Higgs self-coupling constants, κ_3 and κ_4 , for the SM production. The limits are placed via a simultaneous scan of the profile likelihood ratio $-\ln \Delta L$ over the two modifiers, while assuming SM values for all other coupling constants. At the 95% confidence level, assuming $\kappa_4 = 1$, κ_3 is constrained between $-11 - 17$, whereas assuming $\kappa_3 = 1$, κ_4 is constrained between $-230 - 240$, for both the expected and observed limits. Additionally, the CLs method is used to place expected (observed) upper limits on the SM HHH signal strength to be 750 (760) times the predicted SM signal strength, corresponding to an observed (expected) upper limit on the cross-section of 59 fb (59 fb) at the 95% confidence level. For the BSM production, the CLs method is again used to place upper limits on the allowed HHH production cross-sections. The expected (observed) upper limits range between 46 – 350 fb (48 – 310 fb) for the TRSM and DM-CPV models, and between 4.7 – 69 fb (5.7 – 38 fb) for production via generic heavy resonances with a narrow decay width, and between 5.2 – 53 fb (6.3 – 39 fb) for ones with a wide decay width.

Aside from the methodology adopted by the $HHH \rightarrow 6b$ search, three alternative methods explored for this search and new physics signatures in general are presented in Chapter 5. They include a strategy inspired by the ATLAS Run-2 search for non-resonant $HH \rightarrow 4b$, which explores a ML-based jet pairing algorithm, analysis regions defined in a two-dimensional plane of the two Higgs boson candidate masses, and the tri-Higgs invariant mass as the main discriminant. Although this alternative approach does not provide a gain in sensitivity compared to the baseline, it prompts important discussions such as the effects of background sculpting and the feasibility of jet pairing with increasing combinatorics. The second alternative method utilises parametrised neural networks, motivated by their ability to smoothly interpolate between physics parameter values such as the mass of a resonance particle. Their advantages and limitations are discussed, where the latter focus on the ambiguity in the choice of parametrisation scheme when applied to data. Several schemes are explored and compared with an unparametrised model. The last alternative method uses kinematic regression. Motivated by its ability to utilise all particles in the final state while maintaining permutation invariance, a Deep Sets neural network is developed. This approach eliminates the need for explicit jet pairing in searches for multi-Higgs production, which becomes steeply challenging as the

number of Higgs bosons increases. Good response and resolution for the 3-momenta of the Higgs boson candidates in $HH \rightarrow 4b$ signal events are achieved. However the performance degrades for the regression of m_{HHH} in $HHH \rightarrow 6b$ signal events as the final state is more complex. Another limitation lies in the prior distribution of the kinematic variable used in the training, where a desired flat distribution is not always possible for resonance searches. However, kinematic regression could still be applicable to other analyses facing challenges such as increasing combinatorics, large missing energy and high multiplicity in the final state.

Lastly, a novel approach to search for new physics using dataset-wide Graph Neural Networks (GNNs) is presented in Chapter 6. The method is motivated by the additional topological features that are revealed through a singular graph representation of an entire dataset, which are unexploited by conventional ML methods. Events are connected based on their similarity in a graph domain, quantified by a distance metric, to form a dataset-wide graph which is then utilised by a GNN. The input variables associated with each event are then aggregated with its connections via a convolution function in each hidden layer of the GNN. This step, known as message passing, injects the topological information encoded in the graph representation, thus improving the discriminating power of the input variables. A case study searching for a leptoquark (LQ) signal with a final state of a pair of b -quarks, a pair of oppositely-charged leptons and E_T^{miss} is conducted to test the feasibility of this novel approach. The associated dominant SM background processes are $t\bar{t}$ and single top productions. Several approaches of graph construction are explored by comparing different types graph domain and distance metrics. For the graph domain, the trade-off between the improved performance achieved by using expertly engineered high-level kinematic variables and the generalisability offered by low-level kinematic variables is discussed. A dedicated clustering algorithm, *EmbeddingNet*, is also developed to construct a latent space, where same-class events form well-separated clusters. Within these graph domains, events in a dataset are connected according to their event similarity quantified by the Euclidean, cosine and Earth Mover's Distance (EMD). Whilst the Euclidean and cosine distances are found to give similar GNN performances, the EMD generally yields better GNN model performance as well as signal sensitivity compared to the baseline DNNs. This is attributed to the EMD being able to more efficiently capture the topological

differences between signal and background events, such that background events are more clustered whilst signal events are more sparsely distributed in the low-level kinematic space. Two different convolution functions are explored in the GNN models, the Graph Convolution Network (GCN) and the Higher-order Graph Neural Network (GraphConv). The combination of GraphConv model and the graph representation constructed using the EMD metric consistently outperforms the baseline DNNs, in both cases where high-level or low-level kinematic variables are used as the model input variables. This indicates that the topologies where signal events are more sparse and background events are more clustered, improve signal background discrimination. This study presents the first attempt of using a dataset-wide graph approach that could improve BSM searches that face low sensitivity achieved by traditional ML approaches.

The main analysis in this thesis, the search for $HHH \rightarrow 6b$, pioneers the first search of its kind at the LHC and places experimental constraints on the trilinear and the quartic Higgs self-coupling modifiers simultaneously for the first time. Albeit that the range of limits on κ_4 is large and the limits on κ_3 are far from competitive compared to ones placed by HH analyses, this search acts as a foundation for future analyses. The algorithmic development in the ATLAS b -jet triggers will also provide better signal purity from the data-taking perspective, crucial for statistically limited analyses searching for multi-Higgs productions. The alternative strategies as well as the dataset-wide GNN method explored in this thesis provide interesting discussions and fresh perspectives for BSM searches in the future. With the growing dataset, better triggering and the ever-evolving analysis strategies, more stringent constraints on the Higgs self-coupling will continue to be placed. They will lead to more precise determination of the shape of the Higgs potential, the nature of the electroweak symmetry breaking as well as the vacuum stability of the Universe, all of which are important open questions to be answered. The searches for BSM physics will exist in parallel, as the wealth of theories that provide feasible answers to these questions should continue to be explored. They all point to an exciting prospect for the field of particle physics research well into the future.

References

- [1] ATLAS Collaboration, *Search for Triple Higgs Boson Production in the $6b$ Final State using pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, *Phys. Rev. D*, vol. 111, p. 032 006, 3 Feb. 2025. DOI: [10.1103/PhysRevD.111.032006](https://doi.org/10.1103/PhysRevD.111.032006).
- [2] *Configuration, Performance, and Commissioning of the ATLAS b -jet Triggers for the 2022 and 2023 LHC data-taking periods*, *Journal of Instrumentation*, vol. 20, no. 03, P03002, Mar. 2025. DOI: [10.1088/1748-0221/20/03/P03002](https://doi.org/10.1088/1748-0221/20/03/P03002).
- [3] ATLAS Collaboration, *Expected Performance of the ATLAS b -jet Trigger Algorithms for the LHC Run 3*, 2023. [Online]. Available: https://twiki.cern.ch/twiki/bin/view/AtlasPublic/BJetTriggerPublicResults#Expected_Performance_of_the_AN1.
- [4] ATLAS Collaboation, *Standard Model Summary Plots October 2023*, 2023. [Online]. Available: <https://cds.cern.ch/record/2882448>.
- [5] D. E. Morrissey and M. J. Ramsey-Musolf, *Electroweak Baryogenesis*, *New Journal of Physics*, vol. 14, no. 12, p. 125 003, Dec. 2012. DOI: [10.1088/1367-2630/14/12/125003](https://doi.org/10.1088/1367-2630/14/12/125003). [Online]. Available: <https://dx.doi.org/10.1088/1367-2630/14/12/125003>.
- [6] M. Reichert, A. Eichhorn, H. Gies, J. M. Pawlowski, T. Plehn, and M. M. Scherer, *Probing Baryogenesis Through the Higgs Boson Self-coupling*, *Physical Review D*, vol. 97, no. 7, Apr. 2018, ISSN: 2470-0029. DOI: [10.1103/physrevd.97.075008](https://doi.org/10.1103/physrevd.97.075008).
- [7] J. Kinnunen, J. Baarsma, J.-P. Martikainen, and P. Törmä, *The Fulde-Ferrel-Larkin-Ovchinnikov State for Ultracold Fermions in Lattice and Harmonic Potentials*, Jun. 2017. DOI: [10.48550/arXiv.1706.07076](https://doi.org/10.48550/arXiv.1706.07076).
- [8] C. Caprini, R. Durrer, and G. Servant, *Gravitational Wave Generation from Bubble Collisions in First-order Phase Transitions: An Analytic Approach*, Jun. 2008. DOI: [10.1103/physrevd.77.124015](https://doi.org/10.1103/physrevd.77.124015).
- [9] ATLAS Collaboration, *Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC*, *Physics Letters B*, vol. 716, no. 1, pp. 1–29, 2012, ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.020>.

- [10] CMS Collaboration, *Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC*, *Physics Letters B*, vol. 716, no. 1, pp. 30–61, 2012, ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2012.08.021>.
- [11] M. Grazzini, G. Heinrich, S. Jones, et al., *Higgs Boson Pair Production at NNLO with Top Quark Mass Effects*, *Journal of High Energy Physics*, vol. 2018, no. 5, May 2018, ISSN: 1029-8479. DOI: [10.1007/jhep05\(2018\)059](https://doi.org/10.1007/jhep05(2018)059).
- [12] F. A. Dreyer and A. Karlberg, *Vector-boson fusion Higgs pair production at $N^3\text{LO}$* , *Physical Review D*, vol. 98, no. 11, Dec. 2018, ISSN: 2470-0029. DOI: [10.1103/physrevd.98.114016](https://doi.org/10.1103/physrevd.98.114016).
- [13] ATLAS Collaboration, *Search for Non-resonant Pair Production of Higgs Bosons in the $b\bar{b}b\bar{b}$ Final State in pp Collisions at $\sqrt{s}=13\text{ TeV}$ with the ATLAS Detector*, *Phys. Rev. D*, vol. 108, no. 5, DOI: [10.1103/PhysRevD.108.052003](https://doi.org/10.1103/PhysRevD.108.052003). arXiv: [2301.03212 \[hep-ex\]](https://arxiv.org/abs/2301.03212).
- [14] ATLAS Collaboration, *Constraints on the Higgs Boson Self-coupling from Single- and Double-Higgs Production with the ATLAS Detector Using pp Collisions at $\sqrt{s} = 13\text{ TeV}$* , *Physics Letters B*, vol. 843, p. 137 745, Aug. 2023, ISSN: 0370-2693. DOI: [10.1016/j.physletb.2023.137745](https://doi.org/10.1016/j.physletb.2023.137745).
- [15] D. de Florian, I. Fabre, and J. Mazzitelli, *Triple Higgs Production at Hadron Colliders at NNLO in QCD*, *Journal of High Energy Physics*, no. 03, Mar. 2020. DOI: [10.1007/JHEP03\(2020\)155](https://doi.org/10.1007/JHEP03(2020)155).
- [16] J. D. Wells, *Higgs Naturalness and the Scalar Boson Proliferation Instability Problem*, *Synthese*, vol. 194, no. 2, pp. 477–490, Dec. 2014, ISSN: 1573-0964. DOI: [10.1007/s11229-014-0618-8](https://doi.org/10.1007/s11229-014-0618-8).
- [17] L. Canetti, M. Drewes, and M. Shaposhnikov, *Matter and Antimatter in the Universe*, *New Journal of Physics*, vol. 14, no. 9, p. 095 012, Sep. 2012, ISSN: 1367-2630. DOI: [10.1088/1367-2630/14/9/095012](https://doi.org/10.1088/1367-2630/14/9/095012).
- [18] M. Torabian, *Electroweak Vacuum Stability and the Higgs Field Relaxation via Gravitational Effects*, *Classical and Quantum Gravity*, vol. 37, no. 6, p. 065 009, Feb. 2020, ISSN: 1361-6382. DOI: [10.1088/1361-6382/ab6997](https://doi.org/10.1088/1361-6382/ab6997).

- [19] G. Branco, P. Ferreira, L. Lavoura, M. Rebelo, M. Sher, and J. P. Silva, *Theory and Phenomenology of Two-Higgs-Doublet Models*, *Physics Reports*, vol. 516, no. 1-2, pp. 1–102, Jul. 2012, ISSN: 0370-1573. DOI: [10.1016/j.physrep.2012.02.002](https://doi.org/10.1016/j.physrep.2012.02.002).
- [20] C. Csaki, *The Minimal Supersymmetric Standard Model*, *Modern Physics Letters A*, vol. 11, no. 08, pp. 599–613, Mar. 1996, ISSN: 1793-6632. DOI: [10.1142/s021773239600062x](https://doi.org/10.1142/s021773239600062x).
- [21] A. V. Semenov, M. N. Dubinin, *Triple and Quartic Interactions of Higgs Bosons in the Two-Higgs-doublet model with CP-violation*, *The European Physical Journal C - Particles and Fields*, vol. 28, May 2003, ISSN: 1434-6052. DOI: [10.1140/epjc/s2003-01141-5](https://doi.org/10.1140/epjc/s2003-01141-5).
- [22] L. Fromme, S. J. Huber, and M. Seniuch, *Baryogenesis in the Two-Higgs Doublet Model*, *Journal of High Energy Physics*, vol. 2006, no. 11, p. 038, Nov. 2006, ISSN: 1029-8479. DOI: [10.1088/1126-6708/2006/11/038](https://doi.org/10.1088/1126-6708/2006/11/038).
- [23] A. Beniwal, M. Lewicki, M. White, and A. G. Williams, *Gravitational Waves and Electroweak Baryogenesis in a Global Study of the Extended Scalar Singlet Model*, *Journal of High Energy Physics*, vol. 183, no. 2, Feb. 2019, ISSN: 1029-8479. DOI: [10.1007/JHEP02\(2019\)183](https://doi.org/10.1007/JHEP02(2019)183).
- [24] T. Robens, *Two-Real-Singlet-Model Benchmark Planes*, *Symmetry*, vol. 15, no. 1, p. 27, Dec. 2022, ISSN: 2073-8994. DOI: [10.3390/sym15010027](https://doi.org/10.3390/sym15010027).
- [25] T. Robens, T. Stefaniak, and J. Wittbrodt, *Two-real-scalar-singlet Extension of the SM: LHC Phenomenology and Benchmark Scenarios*, *The European Physical Journal C*, vol. 80, no. 2, Feb. 2020, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-020-7655-x](https://doi.org/10.1140/epjc/s10052-020-7655-x).
- [26] E. Mobs, *The CERN Accelerator Complex - August 2018*, 2018. [Online]. Available: <https://cds.cern.ch/record/2636343/>.
- [27] J. Pequeno, *Computer Generated Image of the Whole ATLAS Detector*, 2008. [Online]. Available: <https://cds.cern.ch/record/1095924>.
- [28] I. Neutelings, *CMS Coordinate System*, [Online]. Available: https://tikz.net/axis3d_cms/.
- [29] ATLAS Collaboration, *Production and Integration of the ATLAS Insertable B-Layer*, *Journal of Instrumentation*, vol. 13, no. 05, T05008, May 2018. DOI: [10.1088/1748-0221/13/05/T05008](https://doi.org/10.1088/1748-0221/13/05/T05008).

- [30] ATLAS Collaboration, *Operation and Performance of the ATLAS Semiconductor Tracker*, *Journal of Instrumentation*, vol. 9, no. 08, P08009, Aug. 2014. DOI: [10.1088/1748-0221/9/08/P08009](https://doi.org/10.1088/1748-0221/9/08/P08009).
- [31] ATLAS Collaboration, *The ATLAS TRT Barrel Detector*, vol. 3, no. 02, P02014, Feb. 2008. DOI: [10.1088/1748-0221/3/02/P02014](https://doi.org/10.1088/1748-0221/3/02/P02014).
- [32] ATLAS Collaboration, *Alignment of the ATLAS Inner Detector in Run 2*, *The European Physical Journal C*, vol. 80, no. 1194, 2020. DOI: <https://doi.org/10.1140/epjc/s10052-020-08700-6>.
- [33] *The ATLAS Experiment at the CERN Large Hadron Collider*, vol. 3, no. 08, S08003, Aug. 2008. DOI: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003).
- [34] B. Stelzer, *The New Small Wheel Upgrade Project of the ATLAS Experiment*, *Nuclear and Particle Physics Proceedings*, vol. 273-275, pp. 1160–1165, 2016, 37th International Conference on High Energy Physics (ICHEP), ISSN: 2405-6014. DOI: <https://doi.org/10.1016/j.nuclphysbps.2015.09.182>.
- [35] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013.
- [36] B. M. Werner Herr, *Concept of Luminosity*, DOI: [10.5170/CERN-2006-002.361](https://doi.org/10.5170/CERN-2006-002.361). [Online]. Available: <https://cds.cern.ch/record/941318/>.
- [37] ATLAS Collaboration, *Luminosity Determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS Detector at the LHC*, *The European Physical Journal C*, vol. 83, no. 10, Oct. 2023, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-11747-w](https://doi.org/10.1140/epjc/s10052-023-11747-w). [Online]. Available: <http://dx.doi.org/10.1140/epjc/s10052-023-11747-w>.
- [38] ATLAS Collaboration, *Public ATLAS Luminosity Results for Run-3 of the LHC*, 2024, <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun3>.
- [39] ATLAS Collaboration, *Operation of the ATLAS Trigger System in Run-2*, *Journal of Instrumentation*, vol. 15, P10004, 2020. DOI: [10.1088/1748-0221/15/10/P10004](https://doi.org/10.1088/1748-0221/15/10/P10004).
- [40] R. Achenbach, P. Adragna, and V. A. et al, *The ATLAS Level-1 Calorimeter Trigger*, *Journal of Instrumentation*, vol. 3, no. 03, P03001, 2008. DOI: [10.1088/1748-0221/3/03/P03001](https://doi.org/10.1088/1748-0221/3/03/P03001).
- [41] *Performance of the ATLAS muon triggers in Run 2*, *Journal of Instrumentation*, vol. 15, no. 09, P09015, Sep. 2020. DOI: [10.1088/1748-0221/15/09/P09015](https://doi.org/10.1088/1748-0221/15/09/P09015).

- [42] ATLAS Collaboration, *Software Performance of the ATLAS Track Reconstruction for LHC Run 3*, *Computing and Software for Big Science*, vol. 8, no. 1, Apr. 2024, ISSN: 2510-2044. DOI: [10.1007/s41781-023-00111-y](https://doi.org/10.1007/s41781-023-00111-y).
- [43] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC*, *The European Physical Journal C*, vol. 77, no. 10, 2017, ISSN: 1434-6052. DOI: [0.1140/epjc/s10052-017-5225-7](https://doi.org/10.1140/epjc/s10052-017-5225-7).
- [44] R. Fruhwirth, *Application of Kalman Filtering to Track and Vertex Fitting*, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 262, no. 2, pp. 444–450, 1987, ISSN: 0168-9002. DOI: [https://doi.org/10.1016/0168-9002\(87\)90887-4](https://doi.org/10.1016/0168-9002(87)90887-4).
- [45] ATLAS Collaboration, *Development of ATLAS Primary Vertex Reconstruction for LHC Run 3*, 2019. [Online]. Available: <https://cds.cern.ch/record/2670380>.
- [46] ATLAS Collaboration, *Reconstruction of Primary Vertices at the ATLAS Experiment in Run 1 pp Collisions at the LHC*, *The European Physical Journal C*, vol. 77, no. 5, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-4887-5](https://doi.org/10.1140/epjc/s10052-017-4887-5).
- [47] ATLAS Collaboration, *Topological Cell Clustering in the ATLAS Calorimeters and Its Performance in LHC Run 1*, *The European Physical Journal C*, vol. 77, no. 7, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5004-5](https://doi.org/10.1140/epjc/s10052-017-5004-5).
- [48] ATLAS Collaboration, *Jet Reconstruction and Performance using Particle Flow with the ATLAS Detector*, *The European Physical Journal C*, vol. 77, no. 7, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-017-5031-2](https://doi.org/10.1140/epjc/s10052-017-5031-2).
- [49] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti-kt Jet Clustering Algorithm*, *Journal of High Energy Physics*, vol. 2008, no. 04, p. 063, Apr. 2008. DOI: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063).
- [50] S. D. Ellis and D. E. Soper, *Successive Combination Jet Algorithm for Hadron Collisions*, *Phys. Rev. D*, vol. 48, pp. 3160–3166, 7 Oct. 1993. DOI: [10.1103/PhysRevD.48.3160](https://doi.org/10.1103/PhysRevD.48.3160).
- [51] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *Better Jet Clustering Algorithms*, *Journal of High Energy Physics*, vol. 1997, no. 08, p. 001, Sep. 1997. DOI: [10.1088/1126-6708/1997/08/001](https://doi.org/10.1088/1126-6708/1997/08/001).

- [52] ATLAS Collaboration, *Performance of Pile-up Mitigation Techniques for Jets in pp Collisions at $\sqrt{s} = 8$ TeV using the ATLAS Detector*, *The European Physical Journal C*, vol. 76, no. 11, 2016, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-016-4395-z](https://doi.org/10.1140/epjc/s10052-016-4395-z).
- [53] L.-T. W. David Krohn Jesse Thaler, *Jet Trimming*, *Journal of High Energy Physics*, vol. 2010, no. 84, 2010, ISSN: 1029-8479. DOI: [10.1007/JHEP02\(2010\)084](https://doi.org/10.1007/JHEP02(2010)084).
- [54] L.-T. W. David Krohn Jesse Thaler, *Jets with Variable R* , *Journal of High Energy Physics*, vol. 2009, no. 06, p. 059, Jun. 2009. DOI: [10.1088/1126-6708/2009/06/059](https://doi.org/10.1088/1126-6708/2009/06/059).
- [55] ATLAS Collaboration, *Jet Energy Scale and Resolution Measured in Proton-proton Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, *The European Physical Journal C*, vol. 81, no. 8, 2021, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-021-09402-3](https://doi.org/10.1140/epjc/s10052-021-09402-3).
- [56] ATLAS Collaboration, *Determination of Jet Calibration and Energy Resolution in Proton-proton Collisions at $\sqrt{s} = 8$ TeV Using the ATLAS Detector*, 2021. DOI: <https://doi.org/10.1140/epjc/s10052-020-08477-8>. arXiv: [1910.04482](https://arxiv.org/abs/1910.04482) [hep-ex].
- [57] ATLAS Collaboration, *New Techniques for Jet Calibration with the ATLAS Detector*, *The European Physical Journal C*, vol. 83, no. 8, 2023, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-11837-9](https://doi.org/10.1140/epjc/s10052-023-11837-9).
- [58] M. Amerl, *The ATLAS jet trigger for Run 3 of the LHC*, 2023. DOI: [10.1051/epjconf/202429502030](https://doi.org/10.1051/epjconf/202429502030).
- [59] ATLAS Collaboration, *ATLAS b -jet Identification Performance and Efficiency Measurement with $t\bar{t}$ Events in pp Collisions at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C*, vol. 79, no. 11, Nov. 2019, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8).
- [60] ATLAS Collaboration, *Measurement of b -tagging Efficiency of c -jets in $t\bar{t}$ Events Using a Likelihood Approach with the ATLAS Detector*, Mar. 2018. [Online]. Available: <https://cds.cern.ch/record/2306649>.
- [61] ATLAS Collaboration, *Calibration of Light-flavour Jet b -tagging Rates on ATLAS Proton-proton Collision Data at $\sqrt{s} = 13\sim$ TeV*, Apr. 2018. [Online]. Available: <https://cds.cern.ch/record/2314418>.

- [62] ATLAS Collaboration, *A Continuous Calibration of the ATLAS Flavor-tagging Classifiers via Optimal Transportation Maps*, 2024. [Online]. Available: <https://cds.cern.ch/record/2911642>.
- [63] ATLAS Collaboration, *Muon Reconstruction and Identification Efficiency in ATLAS using the Full Run 2 pp Collision Data Set at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C*, vol. 81, no. 7, Jul. 2021, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-021-09233-2](https://doi.org/10.1140/epjc/s10052-021-09233-2).
- [64] ATLAS Collaboration, *Optimisation and Performance Studies of the ATLAS b -tagging Algorithms for the 2017-18 LHC Run*, [Online]. Available: <https://cds.cern.ch/record/2273281>.
- [65] ATLAS Collaboration, *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*, arXiv: [1703.06114](https://arxiv.org/abs/1703.06114) [cs.LG].
- [66] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Póczos, R. Salakhutdinov, and A. Smola, *Deep Sets*, arXiv: [1703.06114](https://arxiv.org/abs/1703.06114) [cs.LG].
- [67] ATLAS Collaboration, *Deep Sets Based Neural Networks for Impact Parameter Flavour Tagging in ATLAS*, [Online]. Available: <https://inspirehep.net/literature/1798845>.
- [68] ATLAS Collaboration, *Topological b -hadron Decay Reconstruction and Identification of b -jets with the JetFitter Package in the ATLAS Experiment at the LHC*, 2018. [Online]. Available: <https://cds.cern.ch/record/2645405>.
- [69] S. Alioli, P. Nason, C. Oleari, and E. Re, *A General Framework for Implementing NLO Calculations in Shower Monte Carlo Programs: the POWHEG BOX*, *Journal of High Energy Physics*, vol. 2010, no. 6, Jun. 2010, ISSN: 1029-8479. DOI: [10.1007/jhep06\(2010\)043](https://doi.org/10.1007/jhep06(2010)043).
- [70] T. Sjöstrand, S. Ask, J. R. Christiansen, *et al.*, *An Introduction to PYTHIA 8.2*, *Computer Physics Communications*, vol. 191, pp. 159–177, Jun. 2015, ISSN: 0010-4655. DOI: [10.1016/j.cpc.2015.01.024](https://doi.org/10.1016/j.cpc.2015.01.024).
- [71] D. J. Lange, *The EvtGen Particle Decay Simulation Package*, *Nucl. Instrum. Meth. A*, vol. 462, S. Erhan, P. Schlein, and Y. Rozen, Eds., pp. 152–155, 2001. DOI: [10.1016/S0168-9002\(01\)00089-4](https://doi.org/10.1016/S0168-9002(01)00089-4).
- [72] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>.

- [73] J. Barr, J. Birk, M. Dragnet, *et al.*, *Umami: A Python Toolkit for Jet Flavour Tagging*, *Journal of Open Source Software*, vol. 9, no. 102, p. 5833, 2024. DOI: [10.21105/joss.05833](https://doi.org/10.21105/joss.05833).
- [74] ATLAS Collaboration, *Trigger Monitoring and Rate Predictions using Enhanced Bias Data from the ATLAS Detector at the LHC*, [Online]. Available: <https://cds.cern.ch/record/2223498>.
- [75] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, *Attention Is All You Need*, arXiv: [1706.03762](https://arxiv.org/abs/1706.03762) [cs.CL].
- [76] H. Abouabid, A. Arhrib, H. Arnold, *et al.*, *HHH Whitepaper*, *The European Physical Journal C*, vol. 84, no. 11, Nov. 2024, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-024-13376-3](https://doi.org/10.1140/epjc/s10052-024-13376-3).
- [77] A. Papaefstathiou, T. Robens, and G. Tetlalmatzi-Xolocotzi, *Triple Higgs Boson Production at the Large Hadron Collider with Two Real Singlet Scalars*, *Journal of High Energy Physics*, vol. 2021, no. 5, May 2021, ISSN: 1029-8479. DOI: [10.1007/jhep05\(2021\)193](https://doi.org/10.1007/jhep05(2021)193).
- [78] T.-K. Chen, C.-W. Chiang, and I. Low, *Simple Model of Dark Matter and CP Violation*, *Physical Review D*, vol. 105, no. 7, Apr. 2022, ISSN: 2470-0029. DOI: [10.1103/physrevd.105.075025](https://doi.org/10.1103/physrevd.105.075025).
- [79] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5: Going Beyond*, *Journal of High Energy Physics*, vol. 2011, no. 6, Jun. 2011, ISSN: 1029-8479. DOI: [10.1007/jhep06\(2011\)128](https://doi.org/10.1007/jhep06(2011)128).
- [80] R. D. Ball, V. Bertone, S. Carrazza, *et al.*, *Parton Distributions for the LHC Run II*, *Journal of High Energy Physics*, vol. 2015, no. 4, Apr. 2015, ISSN: 1029-8479. DOI: [10.1007/jhep04\(2015\)040](https://doi.org/10.1007/jhep04(2015)040).
- [81] *ATLAS Pythia 8 Tunes to 7 TeV Data*, Nov. 2014. [Online]. Available: <https://cds.cern.ch/record/1966419>.
- [82] CERN, *CERN Yellow Reports: Monographs, Vol 2 (2017): Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, 2017. DOI: [10.23731/CYRM-2017-002](https://doi.org/10.23731/CYRM-2017-002).
- [83] O. Mattelaer, *On the Maximal Use of Monte Carlo Samples: Re-weighting Events at NLO Accuracy*, *The European Physical Journal C*, vol. 76, no. 12, Dec. 2016, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-016-4533-7](https://doi.org/10.1140/epjc/s10052-016-4533-7).

- [84] D. de Florian, I. Fabre, and J. Mazzitelli, *Triple Higgs Production at Hadron Colliders at NNLO in QCD*, *Journal of High Energy Physics*, vol. 2020, no. 3, Mar. 2020, ISSN: 1029-8479. DOI: [10.1007/jhep03\(2020\)155](https://doi.org/10.1007/jhep03(2020)155).
- [85] ATLAS Collaboration, *AtFast3: The Next Generation of Fast Simulation in ATLAS*, *Computing and Software for Big Science*, vol. 6, no. 1, Mar. 2022, ISSN: 2510-2044. DOI: [10.1007/s41781-021-00079-7](https://doi.org/10.1007/s41781-021-00079-7).
- [86] T. Sjöstrand, S. Mrenna, and P. Skands, *A Brief Introduction to Pythia 8.1*, *Computer Physics Communications*, vol. 178, no. 11, pp. 852–867, Jun. 2008, ISSN: 0010-4655. DOI: [10.1016/j.cpc.2008.01.036](https://doi.org/10.1016/j.cpc.2008.01.036).
- [87] R. D. Ball, V. Bertone, S. Carrazza, *et al.*, *Parton Distributions with LHC Data*, *Nuclear Physics B*, vol. 867, no. 2, pp. 244–289, Feb. 2013, ISSN: 0550-3213. DOI: [10.1016/j.nuclphysb.2012.10.003](https://doi.org/10.1016/j.nuclphysb.2012.10.003).
- [88] ATLAS Collaboration, *The Pythia 8 A3 Tune Description of ATLAS Minimum Bias and Inelastic Measurements Incorporating the Donnachie-Landshoff Diffractive Model*, Aug. 2016. [Online]. Available: <https://cds.cern.ch/record/2206965>.
- [89] ATLAS Collaboration, *ATLAS Flavour-tagging Algorithms for the LHC Run 2 pp Collision Dataset*, *The European Physical Journal C*, vol. 83, no. 7, Jul. 2023, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-11699-1](https://doi.org/10.1140/epjc/s10052-023-11699-1).
- [90] ATLAS Collaboration, *Evidence for the $H \rightarrow b\bar{b}$ Decay with the ATLAS Detector*, *Journal of High Energy Physics*, vol. 2017, no. 12, Dec. 2017, ISSN: 1029-8479. DOI: [10.1007/jhep12\(2017\)024](https://doi.org/10.1007/jhep12(2017)024).
- [91] B. Rozemberczki, L. Watson, P. Bayer, *et al.*, *The Shapley Value in Machine Learning*, 2022. arXiv: [2202.05594](https://arxiv.org/abs/2202.05594) [cs.LG].
- [92] ATLAS Collaboration, *Luminosity Determination in pp Collisions at $\sqrt{s} = 13$ TeV using the ATLAS Detector at the LHC*, *The European Physical Journal C*, vol. 83, no. 10, 2023, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-11747-w](https://doi.org/10.1140/epjc/s10052-023-11747-w).
- [93] G. Avoni, M. Bruschi, G. Cabras, *et al.*, *The New LUCID-2 Detector for Luminosity Measurement and Monitoring in ATLAS*, *Journal of Instrumentation*, vol. 13, no. 07, P07017, 2018. DOI: [10.1088/1748-0221/13/07/P07017](https://doi.org/10.1088/1748-0221/13/07/P07017).

- [94] ATLAS Collaboration, *ATLAS b -jet Identification Performance and Efficiency Measurement with $t\bar{t}$ Events in pp Collisions at $\sqrt{s} = 13$ TeV*, *The European Physical Journal C*, vol. 79, no. 11, 2019, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8).
- [95] U. S. Qureshi, R. K. Elayavalli, L. Mozarsky, H. Caines, and I. Mooney, *A New Herwig7 Underlying Event Tune: from RHIC to LHC Energies*, 2024. arXiv: [2411.16897](https://arxiv.org/abs/2411.16897) [hep-ph].
- [96] ATLAS Collaboration, *Configuration and Performance of the ATLAS b -jet Triggers in Run 2*, *The European Physical Journal C*, vol. 81, no. 12, 2021, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-021-09775-5](https://doi.org/10.1140/epjc/s10052-021-09775-5).
- [97] G. Cowan, K. Cranmer, E. Gross, and O. Vitells, *Asymptotic Formulae for Likelihood-based Tests of New Physics*, *The European Physical Journal C*, vol. 71, no. 2, 2011. DOI: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0).
- [98] J. Neyman and E. S. Pearson, *IX. On the Problem of the Most Efficient Tests of Statistical Hypotheses*, 1933, ISSN: 2053-9258. DOI: [10.1098](https://doi.org/10.1098).
- [99] P. Stylianou and G. Weiglein, *Constraints on the Trilinear and Quartic Higgs Couplings from Triple Higgs Production at the LHC and Beyond*, *The European Physical Journal C*, vol. 84, no. 4, 2024, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-024-12722-9](https://doi.org/10.1140/epjc/s10052-024-12722-9).
- [100] ATLAS Collaboration, *A Detailed Map of Higgs Boson Interactions by the ATLAS Experiment Ten Years After the Discovery*, *Nature*, vol. 607, no. 7917, pp. 52–59, 2022, ISSN: 1476-4687. DOI: [10.1038/s41586-022-04893-w](https://doi.org/10.1038/s41586-022-04893-w).
- [101] CMS Collaboration, *A Portrait of the Higgs Boson by the CMS Experiment Ten Years After the Discovery*, *Nature*, vol. 607, no. 7917, pp. 60–68, 2022, ISSN: 1476-4687. DOI: [10.1038/s41586-022-04892-x](https://doi.org/10.1038/s41586-022-04892-x).
- [102] L. H. C. S. W. Group, A. David, A. Denner, et al., *LHC HXSWG Interim Recommendations to Explore the Coupling Structure of a Higgs-like Particle*, 2012. arXiv: [1209.0040](https://arxiv.org/abs/1209.0040).
- [103] ATLAS Collaboration, *HL-LHC Prospects for the Measurement of Triple-Higgs Production in the $6b$ Final State at the ATLAS Experiment*, 2025. [Online]. Available: <https://cds.cern.ch/record/2924772>.

- [104] P. Baldi, K. Cranmer, T. Faucett, P. Sadowski, and D. Whiteson, *Parameterized Neural Networks for High-energy Physics*, *The European Physical Journal C*, vol. 76, no. 5, Apr. 2016, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-016-4099-4](https://doi.org/10.1140/epjc/s10052-016-4099-4).
- [105] J. Zhou, G. Cui, S. Hu, *et al.*, *Graph Neural Networks: A Review of Methods and Applications*, *Open AI*, vol. 1, pp. 57–81, 2020, ISSN: 2666-6510. DOI: <https://doi.org/10.1016/j.aiopen.2021.01.001>.
- [106] ATLAS Collaboration, *Observation of Four-top-quark Production in the Multilepton Final State with the ATLAS Detector*, *The European Physical Journal C*, vol. 83, no. 6, Jun. 2023, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-023-11573-0](https://doi.org/10.1140/epjc/s10052-023-11573-0).
- [107] A. Mullin, S. Nicholls, H. Pacey, M. Parker, M. White, and S. Williams, *Does SUSY Have Friends? A New Approach for LHC Event Analysis*, 2021. arXiv: [1912.10625](https://arxiv.org/abs/1912.10625) [hep-ph].
- [108] T. N. Kipf and M. Welling, *Semi-Supervised Classification with Graph Convolutional Networks*, 2017. arXiv: [1609.02907](https://arxiv.org/abs/1609.02907) [cs.LG].
- [109] C. Morris, M. Ritzert, M. Fey, *et al.*, *Weisfeiler and Leman Go Neural: Higher-order Graph Neural Networks*, 2021. arXiv: [1810.02244](https://arxiv.org/abs/1810.02244) [cs.LG].
- [110] M. Fey and J. E. Lenssen, *Fast Graph Representation Learning with PyTorch Geometric*, 2019. arXiv: [1903.02428](https://arxiv.org/abs/1903.02428) [cs.LG].
- [111] P. T. Komiske, E. M. Metodiev, and J. Thaler, *Metric Space of Collider Events*, *Physical Review Letters*, vol. 123, no. 4, Jul. 2019, ISSN: 1079-7114. DOI: [10.1103/physrevlett.123.041801](https://doi.org/10.1103/physrevlett.123.041801).
- [112] O. Pele and M. Werman, *A Linear Time Histogram Metric for Improved SIFT Matching*, D. Forsyth, P. Torr, and A. Zisserman, Eds., pp. 495–508, 2008.
- [113] B. L. Douglas, *The Weisfeiler-Lehman Method and Graph Isomorphism Testing*, 2011. arXiv: [1101.5211](https://arxiv.org/abs/1101.5211) [math.CO].
- [114] I. Doršner and A. Greljo, *Leptoquark Toolbox for Precision Collider Studies*, *Journal of High Energy Physics*, vol. 2018, no. 5, May 2018, ISSN: 1029-8479. DOI: [10.1007/jhep05\(2018\)126](https://doi.org/10.1007/jhep05(2018)126).

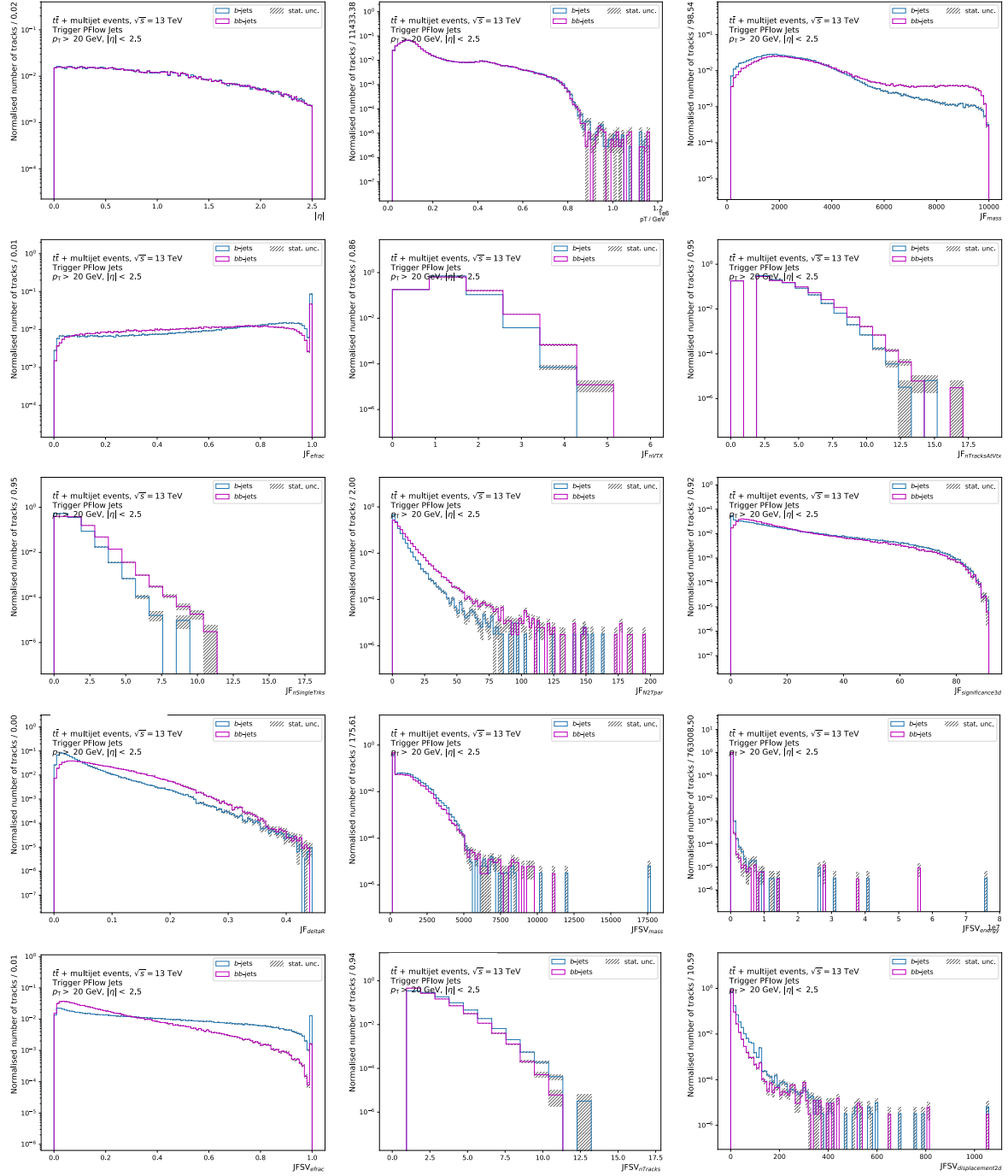
- [115] ATLAS Collaboration, *Combination of Searches for Pair-produced Leptoquarks at $\sqrt{s}=13$ TeV with the ATLAS Detector*, *Physics Letters B*, vol. 854, p. 138736, Jul. 2024, ISSN: 0370-2693. DOI: [10.1016/j.physletb.2024.138736](https://doi.org/10.1016/j.physletb.2024.138736).
- [116] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *MadGraph 5: Going Beyond*, *Journal of High Energy Physics*, vol. 2011, no. 6, Jun. 2011, ISSN: 1029-8479. DOI: [10.1007/jhep06\(2011\)128](https://doi.org/10.1007/jhep06(2011)128).
- [117] C. Bierlich, S. Chakraborty, N. Desai, et al., *A Comprehensive Guide to the Physics and Usage of PYTHIA 8.3*, 2022. arXiv: [2203.11601](https://arxiv.org/abs/2203.11601) [hep-ph].
- [118] J. de Favereau, C. Delaere, P. Demin, et al., *DELPHES 3: a Modular Framework for Fast Simulation of a Generic Collider Experiment*, *Journal of High Energy Physics*, vol. 2014, no. 2, Feb. 2014, ISSN: 1029-8479. DOI: [10.1007/jhep02\(2014\)057](https://doi.org/10.1007/jhep02(2014)057).
- [119] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual: (for version 3.0.2)*, *The European Physical Journal C*, vol. 72, no. 3, Mar. 2012, ISSN: 1434-6052. DOI: [10.1140/epjc/s10052-012-1896-2](https://doi.org/10.1140/epjc/s10052-012-1896-2).
- [120] I. Doršner, S. Fajfer, A. Greljo, J. Kamenik, and N. Košnik, *Physics of Leptoquarks in Precision Experiments and at Particle Colliders*, *Physics Reports*, vol. 641, pp. 1–68, Jun. 2016, ISSN: 0370-1573. DOI: [10.1016/j.physrep.2016.06.001](https://doi.org/10.1016/j.physrep.2016.06.001).
- [121] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-pair Cross-section at Hadron Colliders*, *Computer Physics Communications*, vol. 185, no. 11, pp. 2930–2938, Nov. 2014, ISSN: 0010-4655. DOI: [10.1016/j.cpc.2014.06.021](https://doi.org/10.1016/j.cpc.2014.06.021).
- [122] J. Campbell, T. Neumann, and Z. Sullivan, *Single-top-quark Production in the t-channel at NNLO*, *Journal of High Energy Physics*, vol. 2021, no. 2, Feb. 2021, ISSN: 1029-8479. DOI: [10.1007/jhep02\(2021\)040](https://doi.org/10.1007/jhep02(2021)040).
- [123] N. Kidonakis and N. Yamanaka, *Higher-order Corrections for tW Production at High-energy Hadron Colliders*, *Journal of High Energy Physics*, vol. 2021, no. 5, May 2021, ISSN: 1029-8479. DOI: [10.1007/jhep05\(2021\)278](https://doi.org/10.1007/jhep05(2021)278).
- [124] K. Kumaran, D. J. Papageorgiou, M. Takac, L. Lueg, and N. V. Sahinidis, *Active Metric Learning for Supervised Classification*, *Computers and Chemical Engineering*, vol. 144,

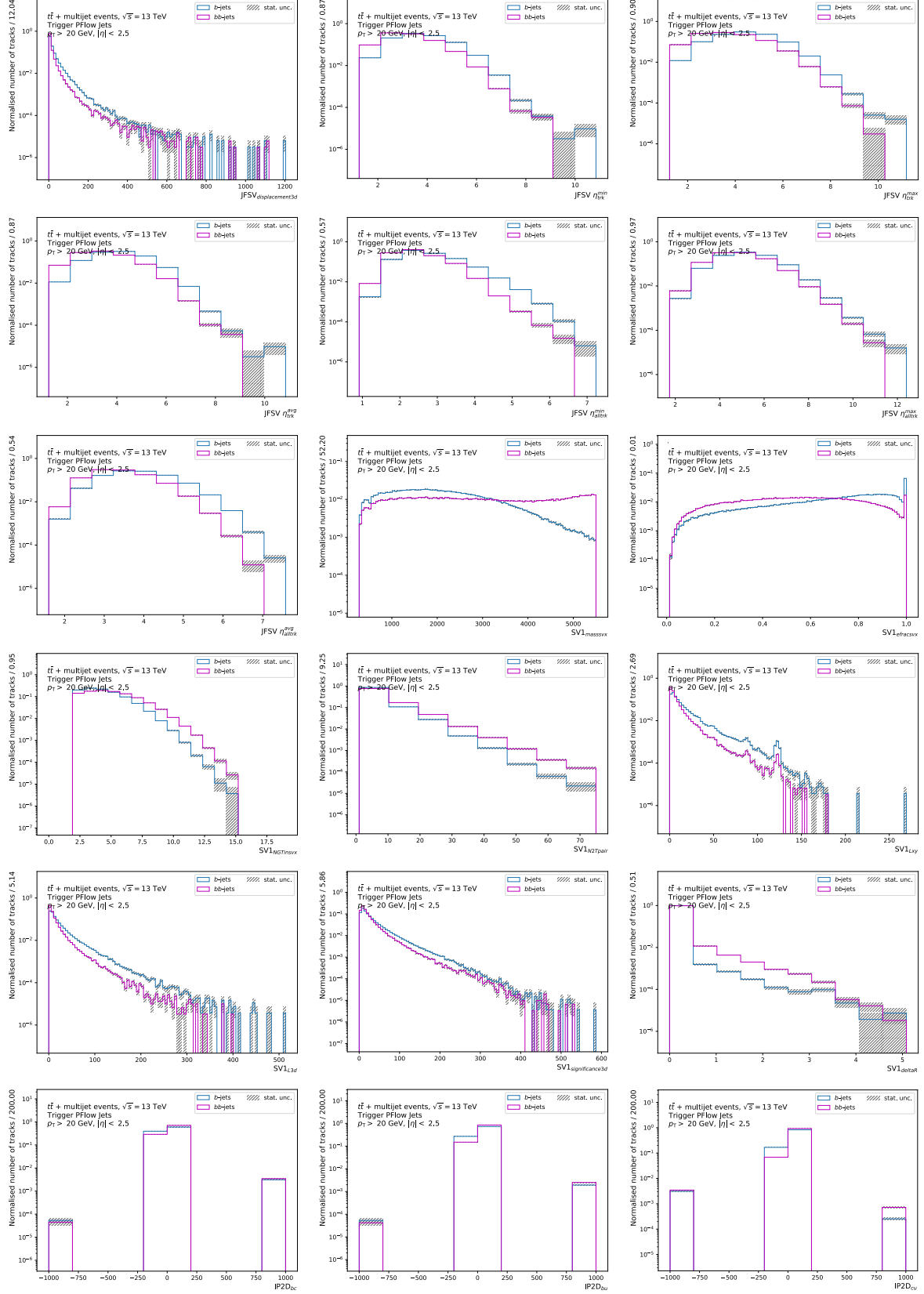
p. 107 132, 2021, ISSN: 0098-1354. DOI: <https://doi.org/10.1016/j.compchemeng.2020.107132>.

- [125] W. L. Hamilton, R. Ying, and J. Leskovec, *Inductive Representation Learning on Large Graphs*, 2018. arXiv: [1706.02216](https://arxiv.org/abs/1706.02216) [cs.SI].
- [126] G. Ciano, A. Rossi, M. Bianchini, and F. Scarselli, *On Inductive? Transductive Learning With Graph Neural Networks*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 2, pp. 758–769, 2022. DOI: [10.1109/TPAMI.2021.3054304](https://doi.org/10.1109/TPAMI.2021.3054304).
- [127] H. Tang and Y. Liu, *Towards Understanding the Generalization of Graph Neural Networks*, 2023. arXiv: [2305.08048](https://arxiv.org/abs/2305.08048) [cs.LG].
- [128] P. Velčković, G. Cucurull, A. Casanova, A. Romero, P. Lió, and Y. Bengio, *Graph Attention Networks*, 2018. arXiv: [1710.10903](https://arxiv.org/abs/1710.10903) [stat.ML].
- [129] B. Xu, N. Wang, T. Chen, and M. Li, *Empirical Evaluation of Rectified Activations in Convolutional Network*, 2015. arXiv: [1505.00853](https://arxiv.org/abs/1505.00853) [cs.LG].

A | DL1dbb Input Variables

This section documents the input variables used to train the DL1dbb algorithm. The distributions are shown for single- b -jets and bb -jets, evaluated on simulated $t\bar{t}$ and QCD multijet samples. The criteria of $p_T > 20$ GeV and $|\eta| < 2.5$ are placed on each jet.





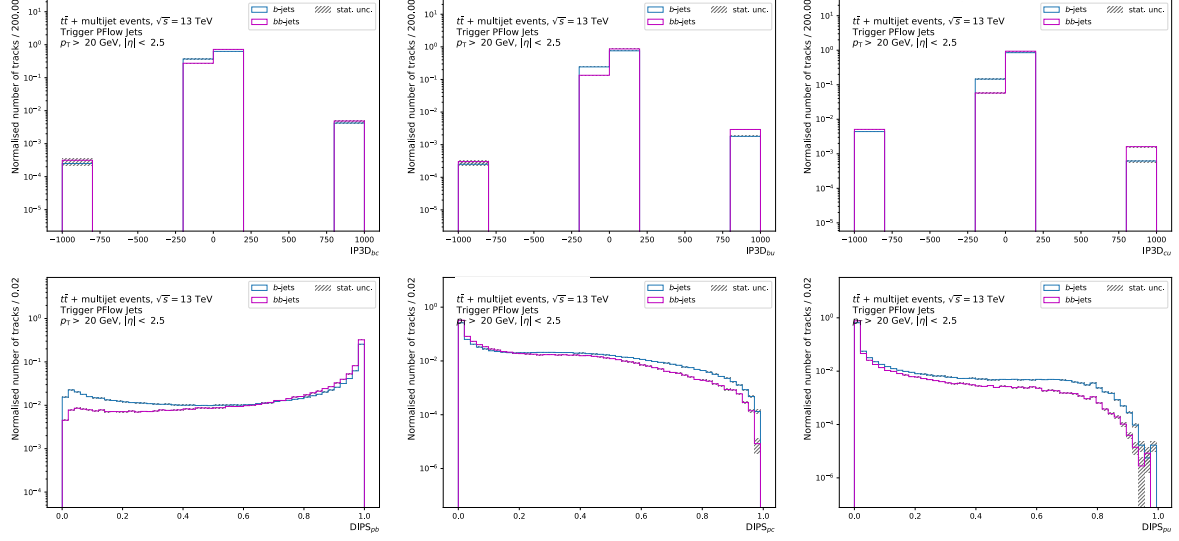


Figure A.1: Input variables used in the training of the DL1dbb summarised in Tables 3.2 and 3.3 of single- b -jets and bb -jets in simulated $t\bar{t}$ and multijet events. All histograms are normalised to have unit area to allow for shape-only comparison.

B | Event Display of a $HHH \rightarrow 6b$ Candidate Event

This section documents the event display generated for a candidate $HHH \rightarrow 6b$ event recorded in the 2018 data-taking year of the ATLAS detector. The event passes the event selections detailed in Section 4.3.3 and is categorised in the $6b$ Signal Region of the analysis. The 6 leading reconstructed b -jets are clearly shown.

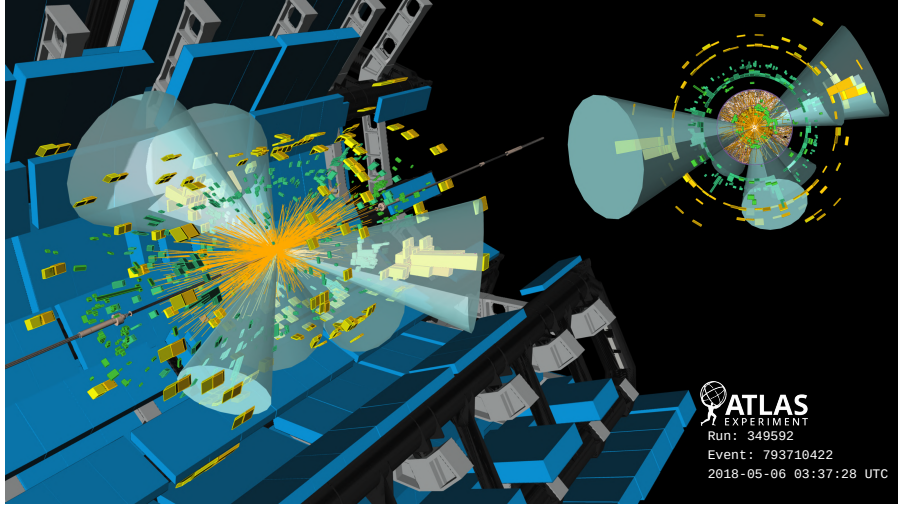


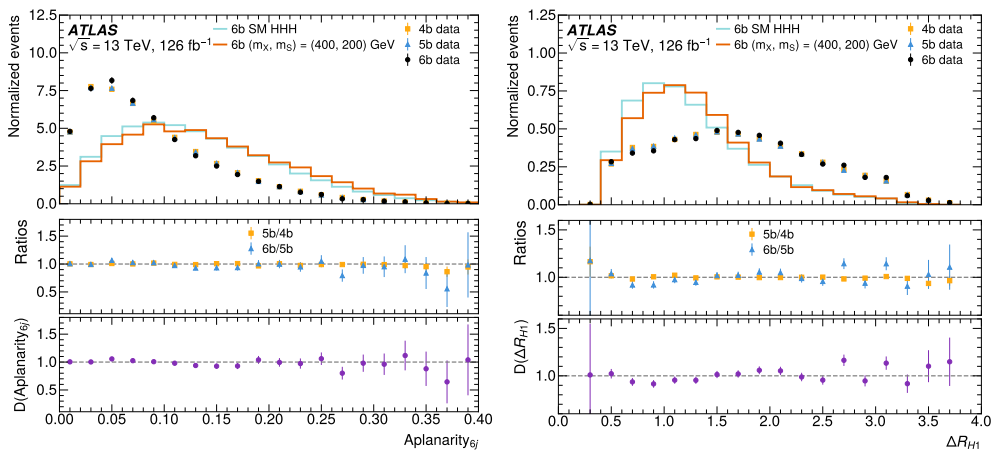
Figure B.1: Event display for a $HHH \rightarrow 6b$ candidate event recorded in 2018, where six jets passing the 77% working point of the b -tagging algorithm DL1d, with $p_T > 40$ GeV and $|\eta| < 2.5$, are shown as blue cones. The particle interactions in the Pixel detector are shown as white dots, in the SCT as yellow strips, and in the TRT as red and white dots. The tracks of charged particles reconstructed in the ID with $p_T > 0.5$ GeV are displayed in orange. The green/teal and yellow/orange cells correspond to the energy deposits in the electromagnetic and hadronic calorimeters, respectively, with $E_T > 250$ MeV. The figure is published in [1].

C | Additional Information on ResDNN, NonresDNN and HeavyresDNN

The kinematic distributions of the input features used in the nonresDNN training are shown in Figure C.1 for the $4b$, $5b$ and $6b$ data, as well as for the $6b$ SM HHH signal sample and a non-resonant TRSM signal sample with $(m_X, m_S) = (400, 200)$ GeV. The matrices of correlations between these input features are shown in Figure C.2 for $5b$ data background and $6b$ SM HHH signal events separately.

The kinematic distributions of the input features used in the resDNN training are shown in Figure C.3 for the $4b$, $5b$ and $6b$ data, as well as for the $6b$ events of two representative resonant TRSM samples with $(m_X, m_S) = (450, 275)$ and $(520, 325)$ GeV. The matrices of correlations between these input features are shown in Figure C.4 for $5b$ data background and combined $6b$ resonant TRSM signal events separately.

The kinematic distributions of the input features used in the heavyresDNN training are shown in Figure C.5 for the $4b$, $5b$ and $6b$ data, as well as for the $6b$ events of two representative heavy-resonant HHH samples with a narrow decay width and $(m_X, m_S) = (700, 400)$ and $(1500, 1000)$ GeV. Figure C.6 show the same kinematics but for the two signal samples with a wide decay width. The matrices of correlations between these input features are shown in Figure C.7 for $5b$ data background and combined narrow and wide width $6b$ heavy-resonant signal events separately.



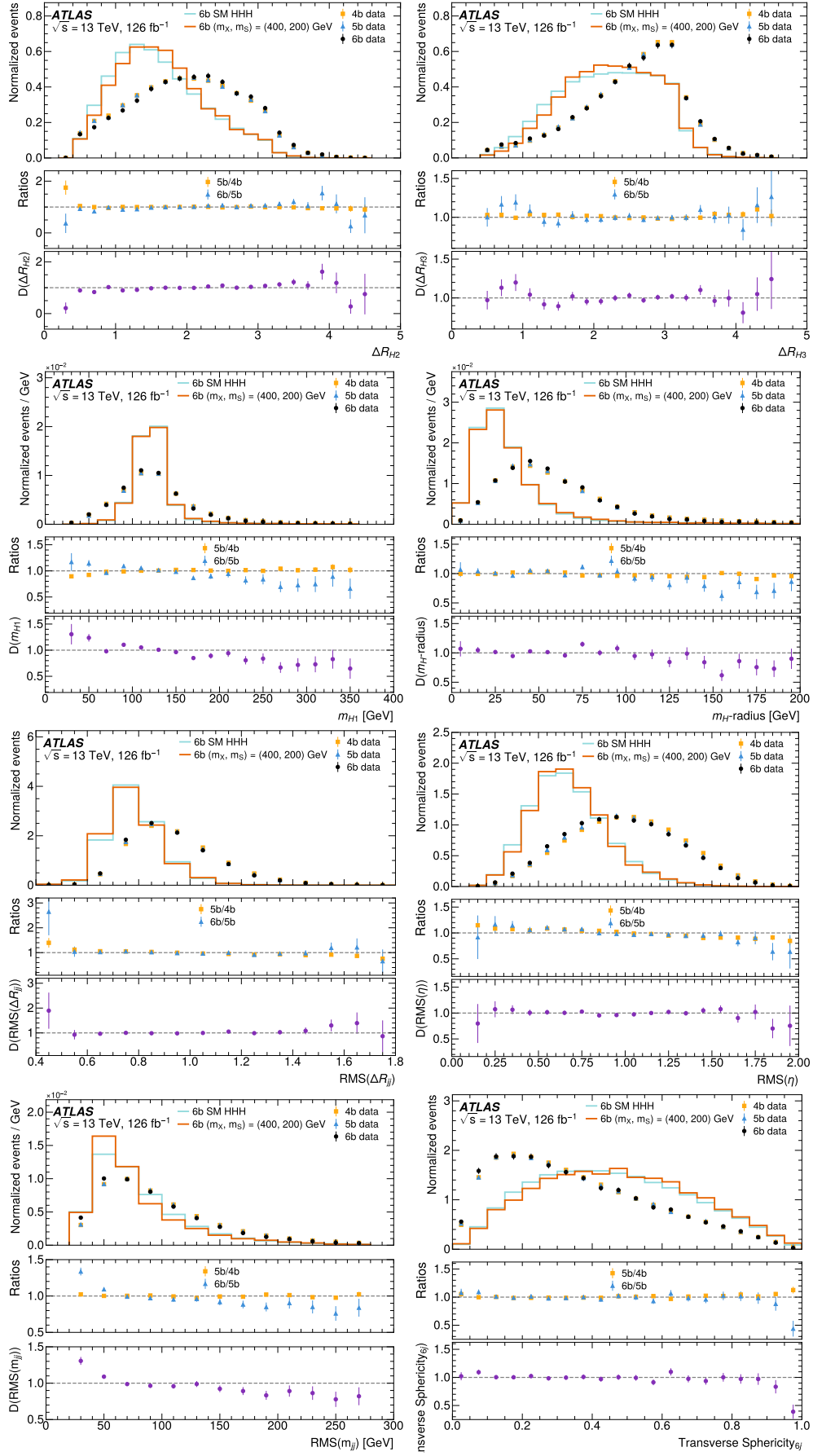


Figure C.1: In the top panel of each plot, kinematic distributions for the nonresDNN. Distributions for $4b$, $5b$ and $6b$ data and example signal samples are shown, with all histograms normalised to unit area. In the middle panel of each plot, ratios of data with different b -jet multiplicities are shown. Finally, in the bottom panel D , the ratio $6b/5b$ divided by the ratio $5b/4b$ is shown. D is used to estimate the size of the systematic uncertainty in the b -jet extrapolation method as described in Section 4.6.3.

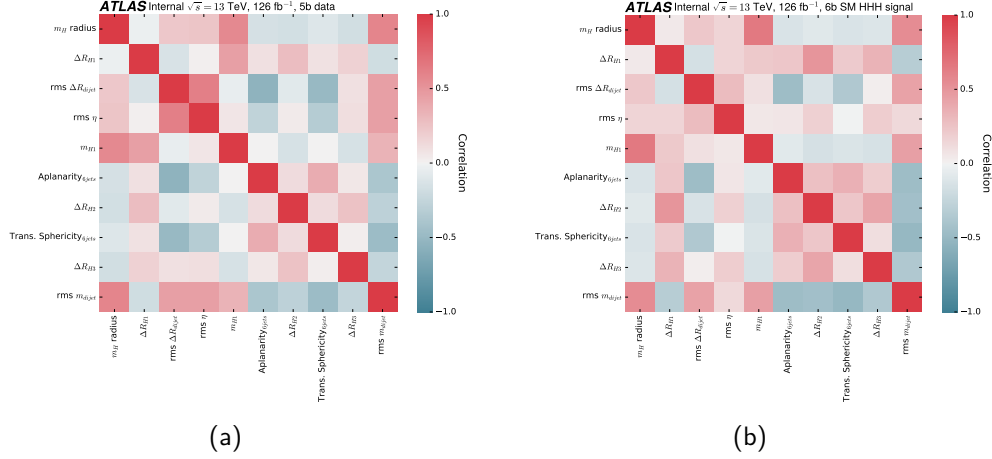
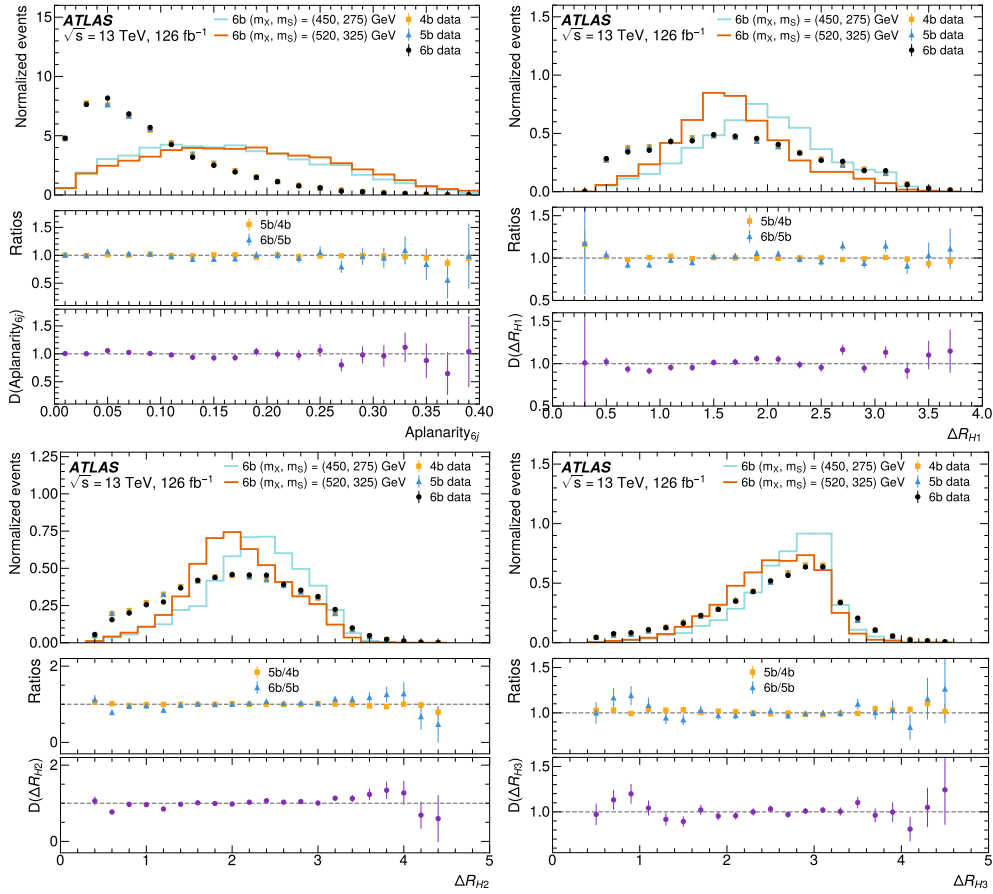


Figure C.2: The matrix of correlation between input features selected for the training of the non-resDNNs, evaluated on the $6b$ SM HHH signal events in (b), and on $5b$ data background in (a).



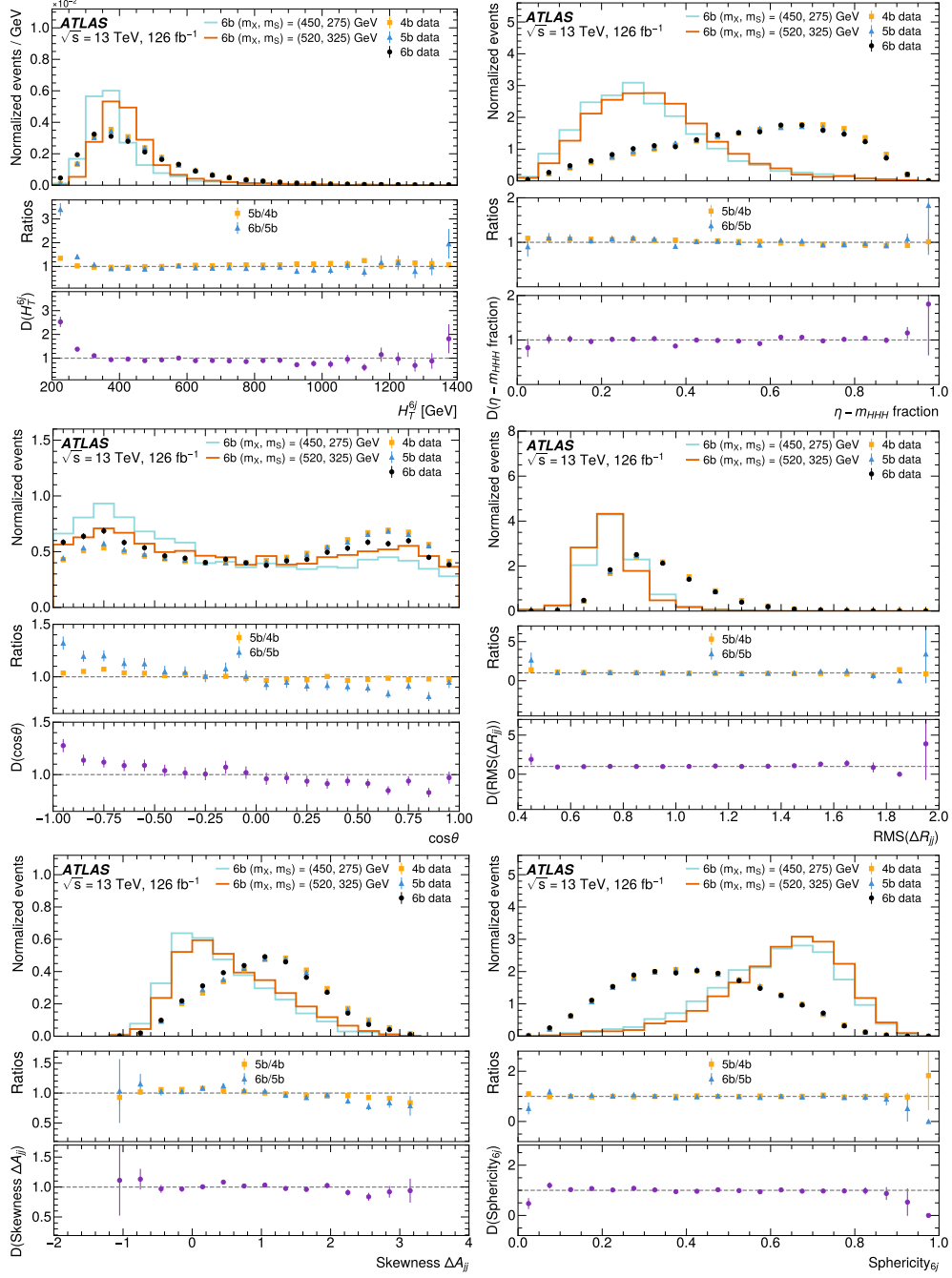


Figure C.3: In the top panel of each plot, kinematic distributions for the resDNN. Distributions for the $4b$, $5b$ and $6b$ data and example signal samples are shown, with all histograms normalised to unit area. In the middle panel of each plot, ratios of data with different b -jet multiplicities are shown. Finally, in the bottom panel D , the ratio $6b/5b$ divided by the ratio $5b/4b$ is shown. D is used to estimate the size of the systematic uncertainty in the b -jet extrapolation method as described in Section 4.6.3.

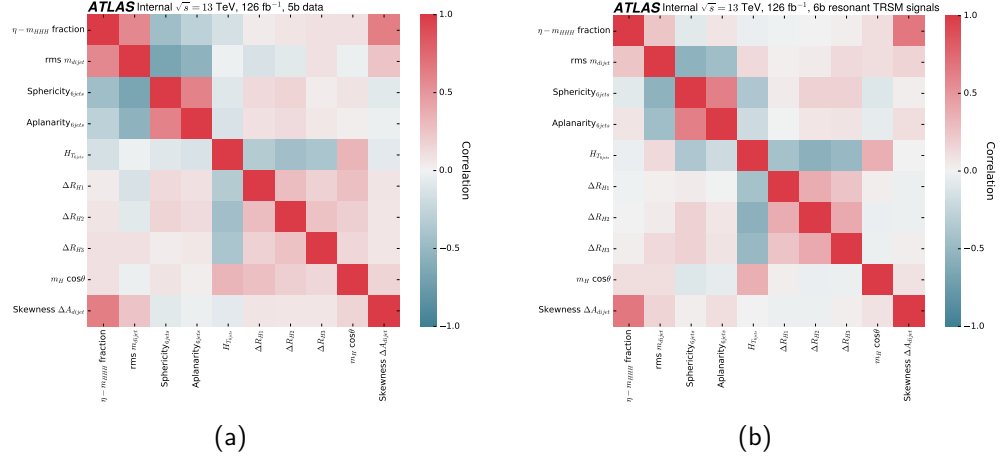
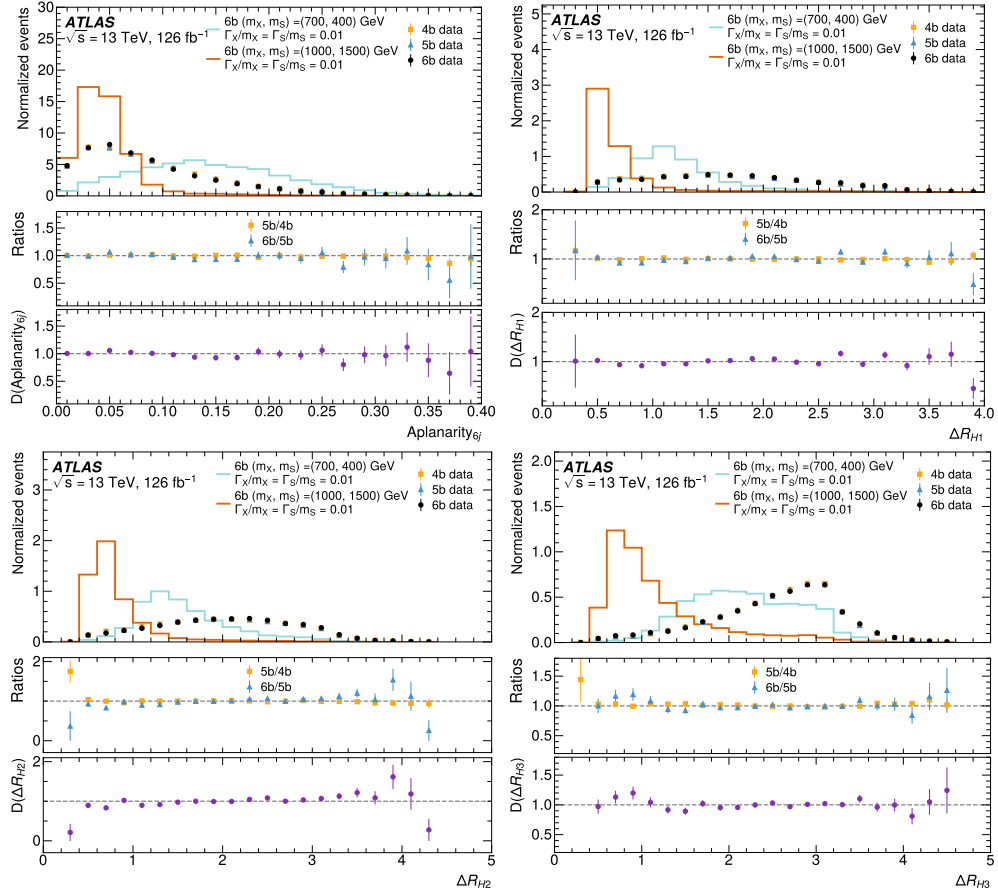


Figure C.4: The matrix of correlation between input features selected for the training of the resDNNs, evaluated on the combined $6b$ resonant TRSM signal events in (b), and on $5b$ data background in (a).



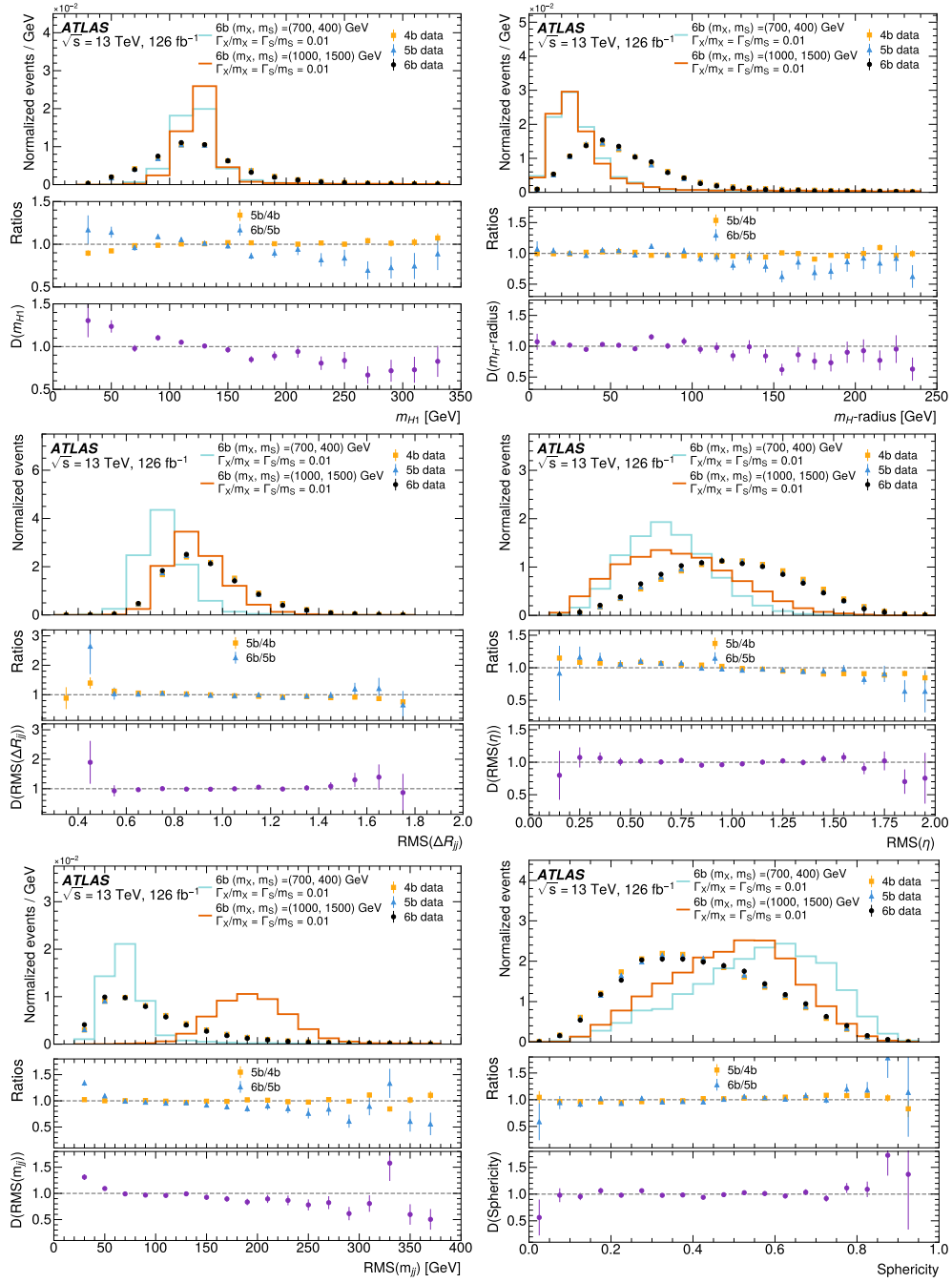
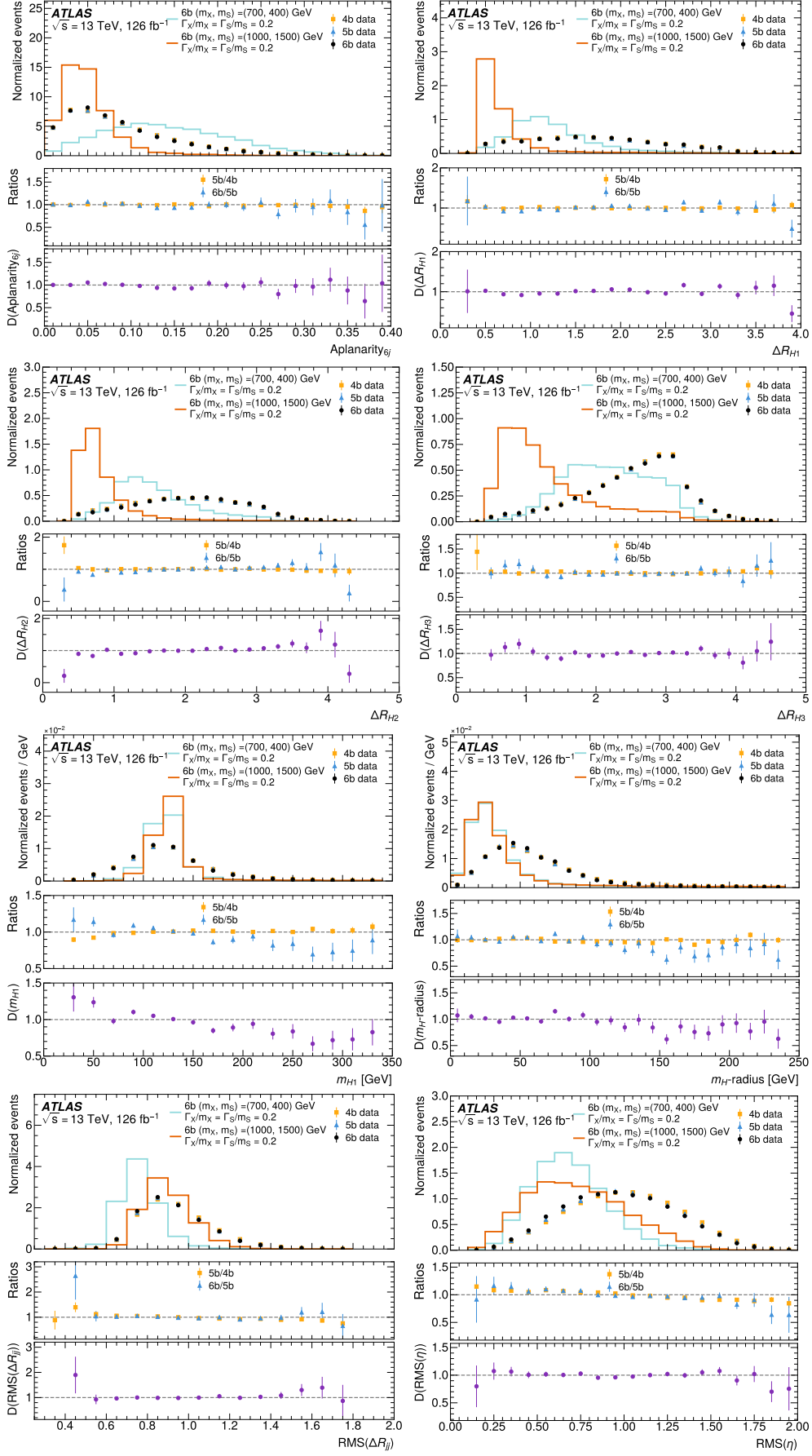


Figure C.5: In the top panel of each plot, kinematic distributions for the heavyresDNN. Distributions for 4b, 5b and 6b data and example narrow width signal samples are shown, with all histograms normalised to unit area. In the middle panel of each plot, ratios of data with different b -jet multiplicities are shown. Finally, in the bottom panel D , the ratio $6b/5b$ divided by the ratio $5b/4b$ is shown. D is used to estimate the size of the systematic uncertainty in the b -jet extrapolation method as described in Section 4.6.3.



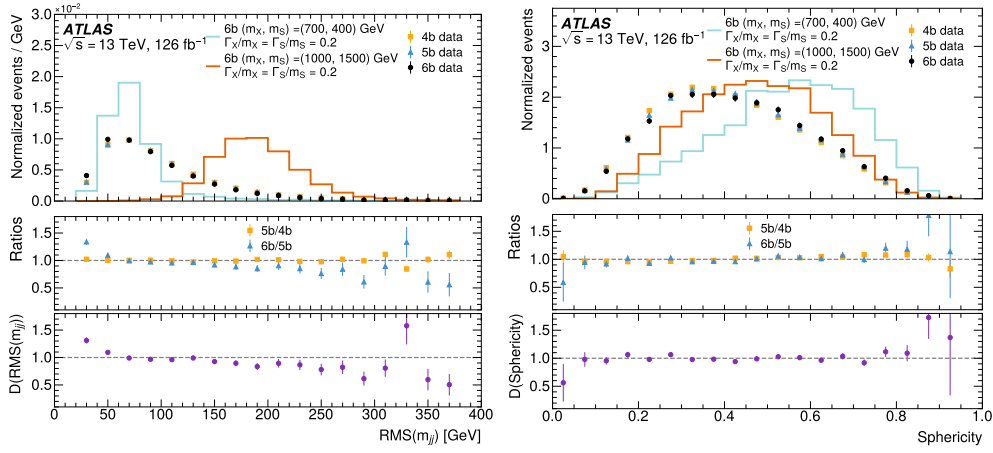


Figure C.6: In the top panel of each plot, kinematic distributions for the heavyresDNN. Distributions for $4b$, $5b$ and $6b$ data and example wide width signal samples are shown, with all histograms normalised to unit area. In the middle panel of each plot, ratios of data with different b -jet multiplicities are shown. Finally, in the bottom panel D , the ratio $6b/5b$ divided by the ratio $5b/4b$ is shown. D is used to estimate the size of the systematic uncertainty in the b -jet extrapolation method as described in Section 4.6.3.

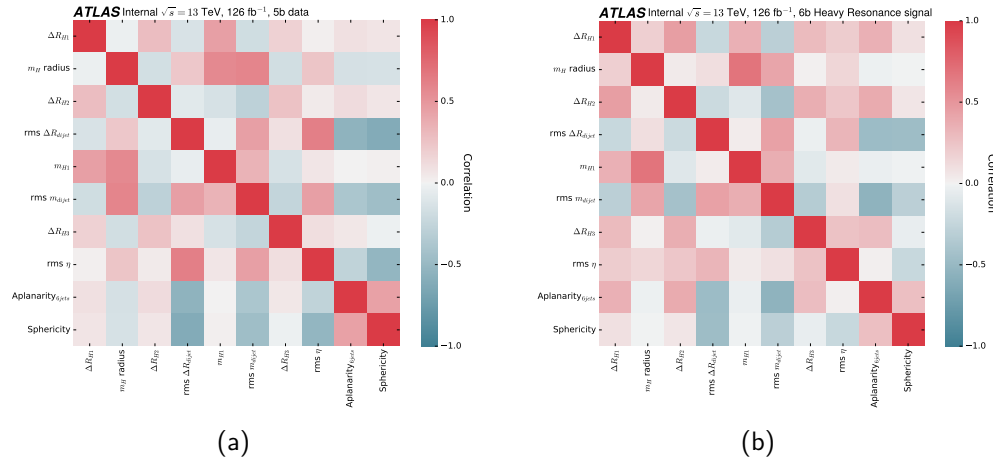


Figure C.7: The matrix of correlation between input features selected for the training of the resDNNs, evaluated on the combined narrow and wide width $6b$ heavy-resonant signal events in (b), and on $5b$ data background in (a).

D | HHH Signal Contamination in Various b -tag Regions

The contamination from the resonant TRSM signal events and HHH signal events are quantified in various b -tag regions by the bin-by-bin Z significance over the DNN scores. The Z significance is calculated for observing n events where n consists of both the $6b$ signal events and the expected number of background events b plus the statistical uncertainty σ on b . The full equation used for this calculation is the following:

$$Z = \sqrt{2 \left(n \ln \left[\frac{n(b + \sigma^2)}{b^2 + n\sigma^2} \right] - \frac{b^2}{\sigma^2} \ln \left[1 + \frac{\sigma^2(n - b)}{b(b + \sigma^2)} \right] \right)} \quad (\text{D.1})$$

The contaminations from $6b$ signal events of 3 representative TRSM scalar masses and the SM signal injected at a nominal cross-section of 50 fb, and 2 representative heavy resonant scalar masses with a narrow and wide decay width injected at a nominal cross-section of 20 fb are shown in Figure D.1. The levels of contamination are checked in the distributions of the associated DNN output scores. The background estimate is derived by scaling down the $5b$ data distribution using the ratio of $\frac{6b \text{ data}}{5b \text{ data}}$. The contaminations are all sufficiently low for the $6b$ data to be used to derive and validate the nominal background estimate, detailed in Section 4.5

The contaminations from $5b$ signal events of the same signal samples as above are checked against the background in the $5b$ region, in this case the unscaled $5b$ data, shown in Figure D.2. This is necessary as a non-negligible amount of contamination is present, indicated by the Z significance especially in the High-score region of the DNN scores. This sufficiently justifies the inclusion of $5b$ signal events in the statistical fitting procedure described in Section 4.8, where the parameter of interest (signal strength) is correlated across both the $6b$ and $5b$ regions.

Lastly, the contaminations from $4b$ signal events of the same signal samples as above are checked against the background in the $4b$ region, which is the unscaled $4b$ data, shown in Figure D.3. The Z significance is low for all signal samples and across the DNN scores. The $4b$ signal events are therefore not taken into consideration in the statistical fitting procedure.

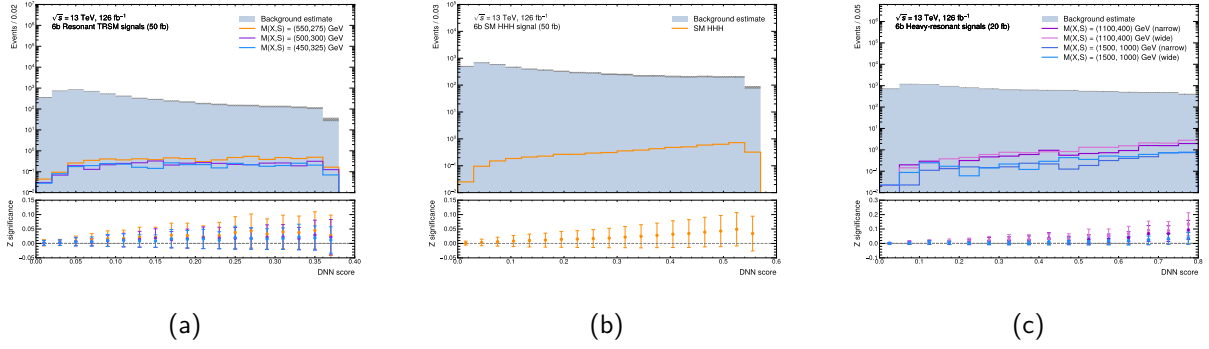


Figure D.1: The output score distributions of the ResDNNs (a), NonresDNNs (b), and the HeavyresDNNs (c), for $6b$ events of several representative signal samples of each interpretation, and for the background estimate in the corresponding Low-score regions of the DNNs. The bottom panel shows the Z significance of each signal sample over the background estimate.

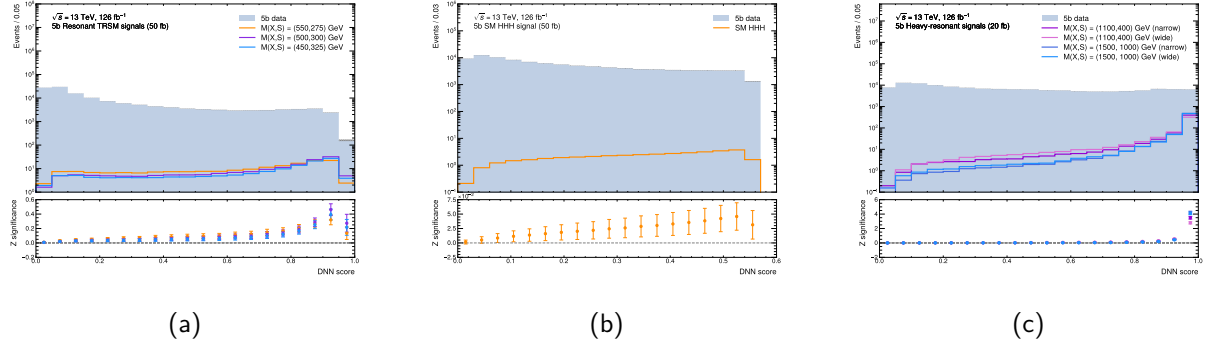


Figure D.2: The output score distributions of the ResDNNs (a), NonresDNNs (b), and the HeavyresDNNs (c), for $5b$ events of several representative signal samples of each interpretation, and for the background estimate in the corresponding Low-score regions of the DNNs. The bottom panel shows the Z significance of each signal sample over the background estimate.

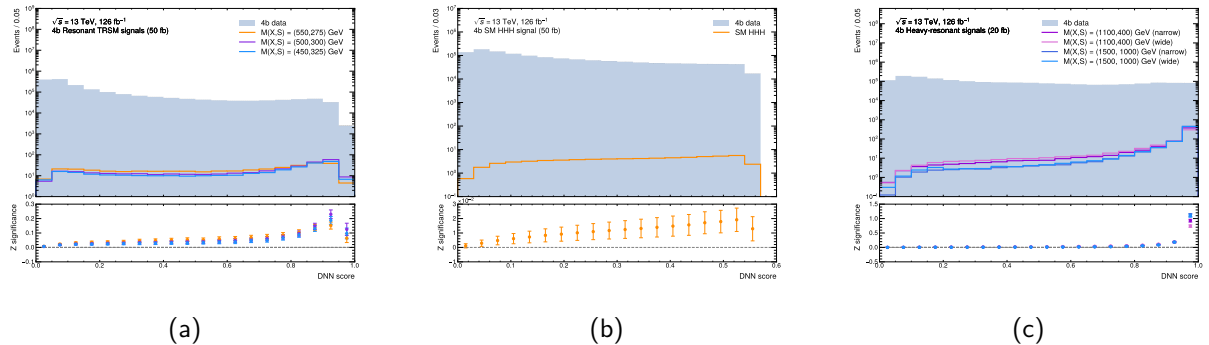


Figure D.3: The output score distributions of the ResDNNs (a), NonresDNNs (b), and the HeavyresDNNs (c), for $4b$ events of several representative signal samples of each interpretation, and for the background estimate in the corresponding Low-score regions of the DNNs. The bottom panel shows the Z significance of each signal sample over the background estimate.

E | Shape Variations of the HHH Background Estimate

This section documents the variations in the shape of the background estimate derived for each interpretation in the $HHH \rightarrow 6b$ analysis. The systematic uncertainties arising from three selected kinematic inputs of the DNN trained for each interpretation are considered. The selection is made such that the variations in the shape of the DNN score are sufficiently described by the variations of the chosen set of kinematic variables. Figures E.1, E.2 and E.3 show the variations in the DNN score of the nominal background estimate in the $6b$ signal region for the non-resonant, resonant and heavy-resonant interpretations respectively.

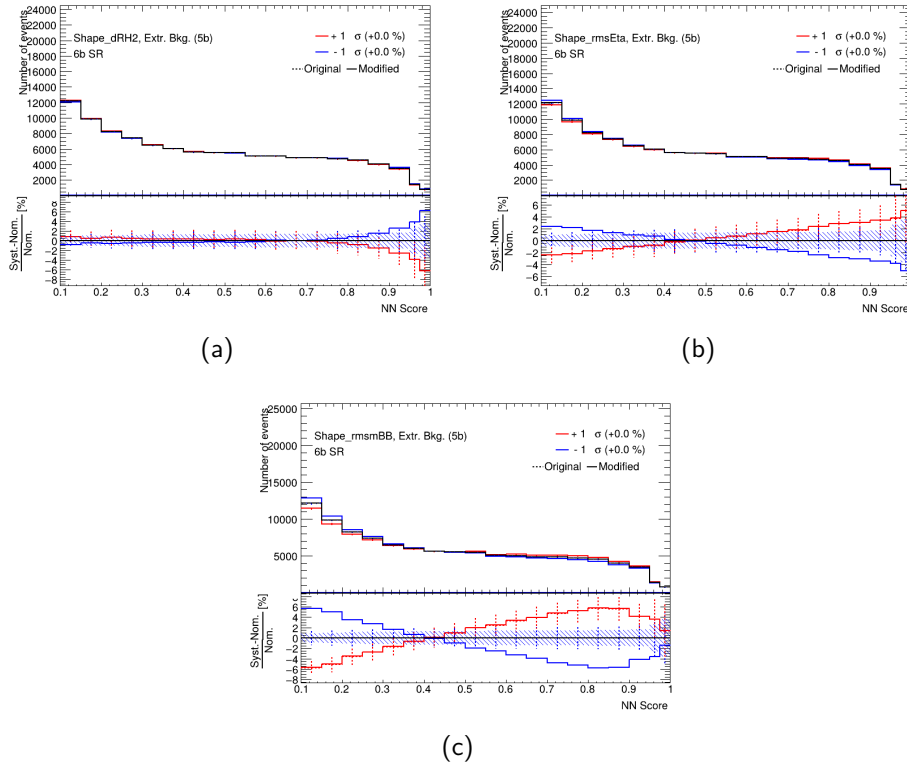


Figure E.1: Up and down variations of the systematic uncertainties on the background estimate for the non-resonant interpretation in the $HHH \rightarrow 6b$ analysis. Image courtesy of Nick Kyriacou.

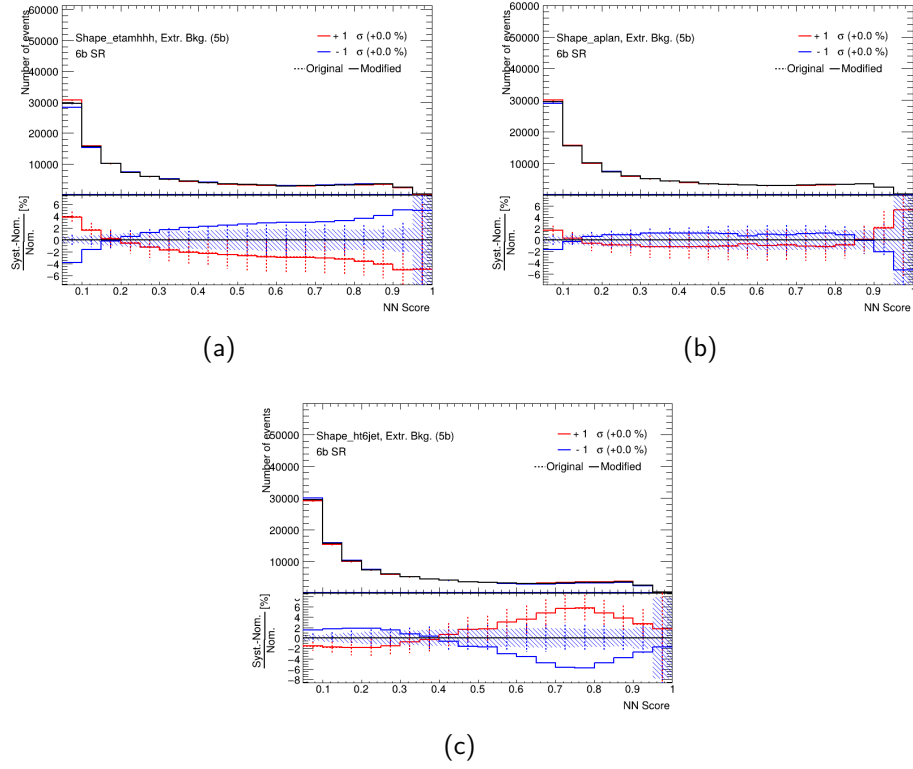


Figure E.2: Up and down variations of the systematic uncertainties on the background estimate for the resonant interpretation in the $HHH \rightarrow 6b$ analysis. Image courtesy of Nick Kyriacou.

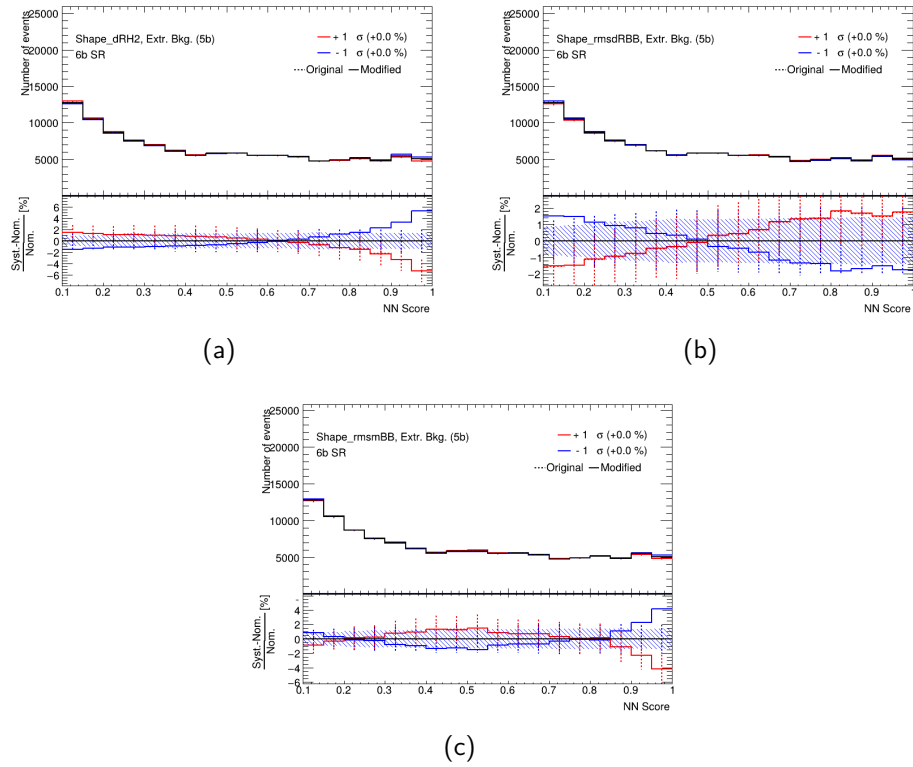


Figure E.3: Up and down variations of the systematic uncertainties on the background estimate for the heavy-resonant interpretation in the $HHH \rightarrow 6b$ analysis. Image courtesy of Nick Kyriacou.

F | EmbeddingNet Algorithm

This section details the training of a machine-learning-based clustering algorithm used to construct the latent space for a dataset-wide graph representation presented in Chapter 6.

F.1 Parameter Scan

A neural network is tasked to perform clustering between same-class events, while providing large separation between the clusters in a high-dimensional latent space. To this end, a contrastive hinge loss is used as the loss function, detailed in Equation 6.14. Within the loss function, the margin m that penalises the separation between pairs of different-class events, and the penalty λ which penalises the separation between different-class clusters are tunable parameters. The dimension of the output latent space, that is the number of embedding features the model outputs per event, can also be tuned. The performance of the algorithm is evaluated based on two metrics:

- Same-class efficiency: $\frac{\text{Edges between same-class nodes}}{\text{Total possible edges}}$.
- Same-class purity: $\frac{\text{Edges between same-class nodes}}{\text{Total edges formed}}$.

where the edges are formed by connecting nodes with a distance smaller than half of the margin m specified in the loss function. Other metrics that are important for the GNN training downstream include edge fraction, which is $\frac{\text{Total edges formed}}{\text{Total possible edges}}$. A scan of the tunable parameters summarised in Table F.1 was performed while the above metrics were evaluated. The values of each parameter in the combination that yields the highest same-class purity in edges formed in the validation dataset are labeled in bold. The correlations between them are shown in Figure F.1.

F.2 Training

The training of the EmbeddingNet models is done using a feedforward deep neural network implemented using PyTorch. The training dataset consists of all possible pairs formed with 400 events randomly sampled from the LQ signal with $m_{LQ} = 1$ TeV, and 400 randomly events sampled from

Parameter	Scanned values
Margin	[0.1 , 0.5, 1.0, 2.0]
Penalty	[1, 5, 10, 20]
Dimension	[2, 6, 12 , 18]

Table F.1: Table summarising the values of the tunable parameters in the contrastive hinge loss used in the EmbeddingNet training that are scanned to optimise the same-class efficiency and purity of the edges formed. The values of the parameters in the combination that gives the highest same-class purity in edges formed in the validation dataset are labeled in bold.

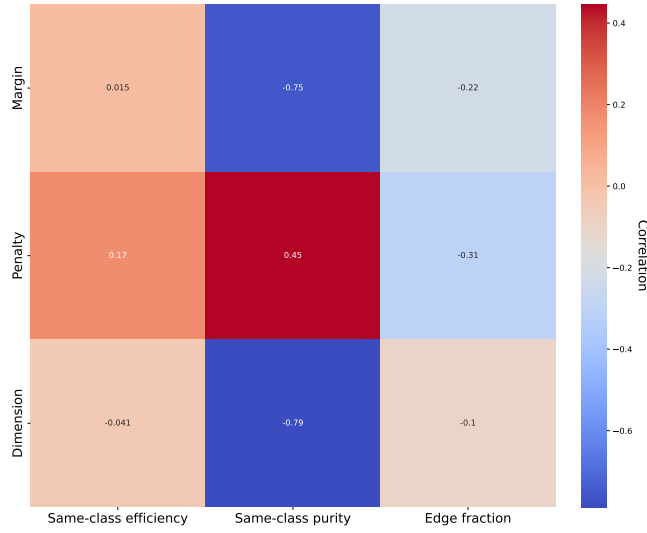


Figure F.1: Matrix of correlations between the tunable parameters in the contrastive hinge loss function of the EmbeddingNet training: margin, penalty and the dimension of the output latent space, and the performance metrics: same-class efficiency and purity, and edge fraction.

the combined SM $t\bar{t}$ and single top backgrounds, passing the $E_T^{\text{miss}} > 200$ GeV requirement. The validation dataset consists of all possible pairs formed with 200 events from each class. Low-level kinematic variables detailed in Section 6.3.2 are used as input features for each event. The model outputs N number of embedding features per event, where N is the dimension of the target latent space.

F.3 Embedding Features in Latent Space

For the model trained using the parameters above, the 12 outputs of the EmbeddingNet model are shown in Figure F.2. They are evaluated on the full dataset of LQ signal events with $m_{LQ} = 1$ TeV and the combined SM $t\bar{t}$ and single top background events used in the GNN model training.

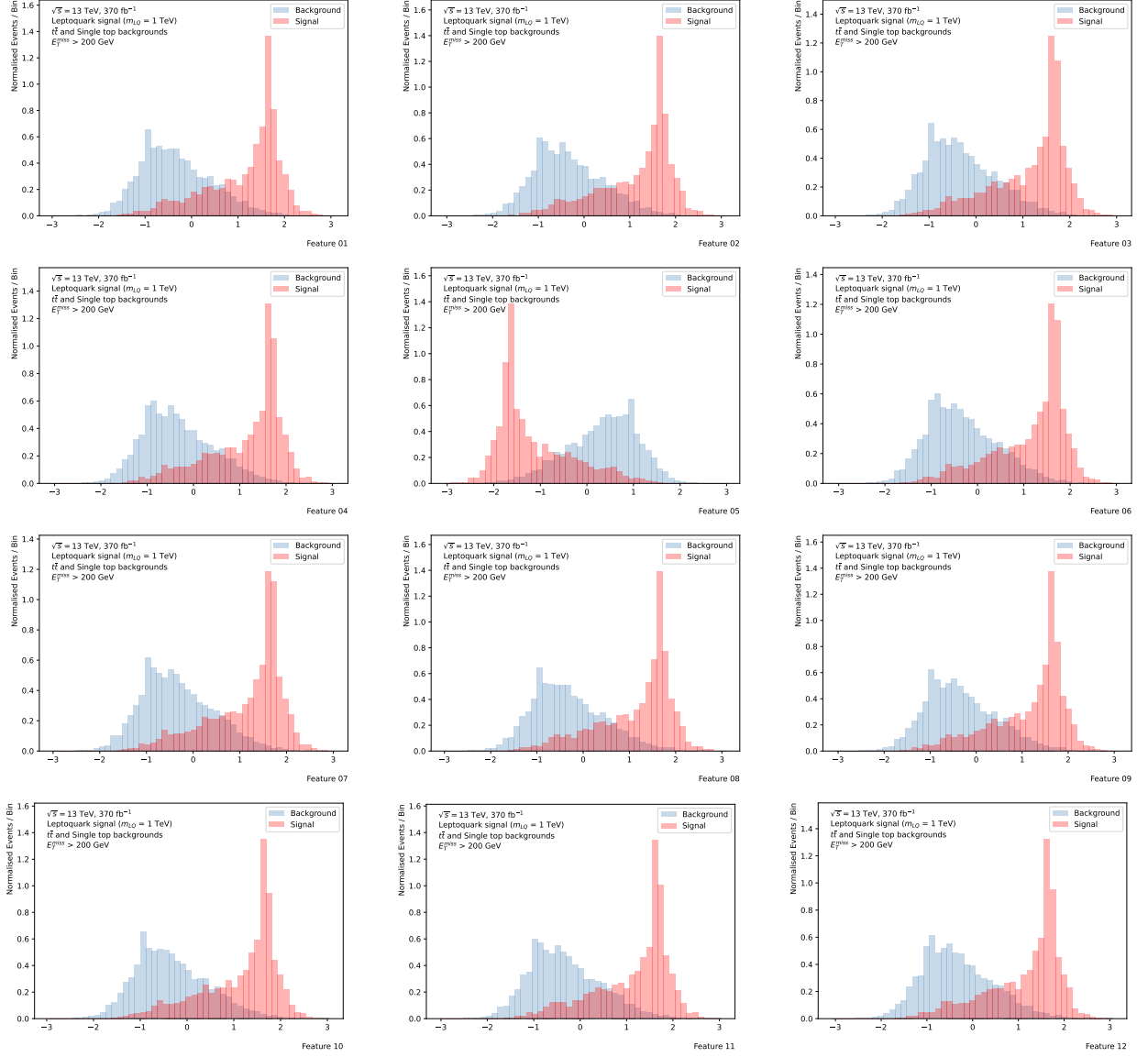


Figure F.2: Distributions of the 12 outputs of the EmbeddingNet model, trained with a margin of 0.1, a penalty of 20 and a latent space dimension of 12. They are shown for the LQ signal events with $m_{LQ} = 1 \text{ TeV}$ and the combined SM $t\bar{t}$ and single top background events used in the GNN model training in Chapter 6. The histograms are normalised to have unit area for shape-only comparison.

G | Additional Information on the Dataset-wide GNNs

This section documents some additional information relating to the dataset-wide GNNs outlined in Chapter 6.

G.1 Kinematic Convolutions

The following shows the distributions of the original, once convoluted and twice convoluted kinematic variables using graph connections made in various graph domain and distance metrics. All kinematic convolutions are calculated using the GCN convolution function. For the high-level kinematic variables, the $p_{T,bbl}$ and $m_{T,l1}$ are shown as two examples. Their convoluted distributions using graphs constructed in the high-level kinematic space, using the Euclidean and cosine distances with a linking length corresponding to a 0.2 edge fraction are shown in Figures G.1 and G.2 respectively. The convoluted distributions using a graph constructed in a 12-dimensional latent space using the Euclidean and cosine distances with a linking length corresponding to a 0.1 edge fraction are shown in Figures G.7 and G.8.

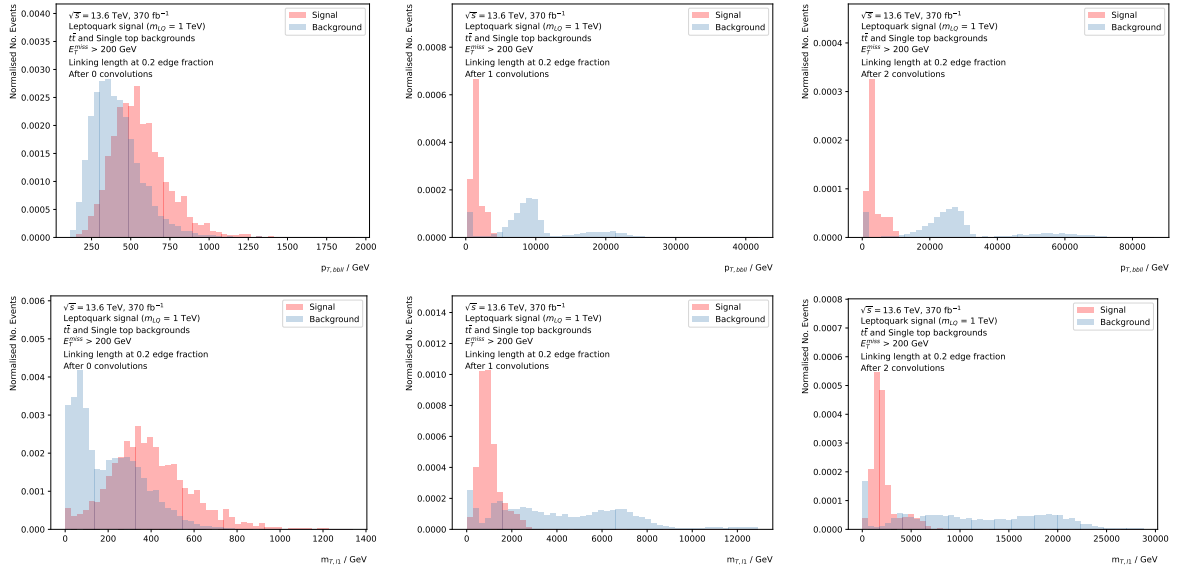


Figure G.1: Distributions of the original, once convoluted and twice convoluted $p_{T,bbl}$ and $m_{T,l1}$ in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in the high-level kinematic domain using the Euclidean distance with a linking length corresponding to a target edge fraction of 0.2. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

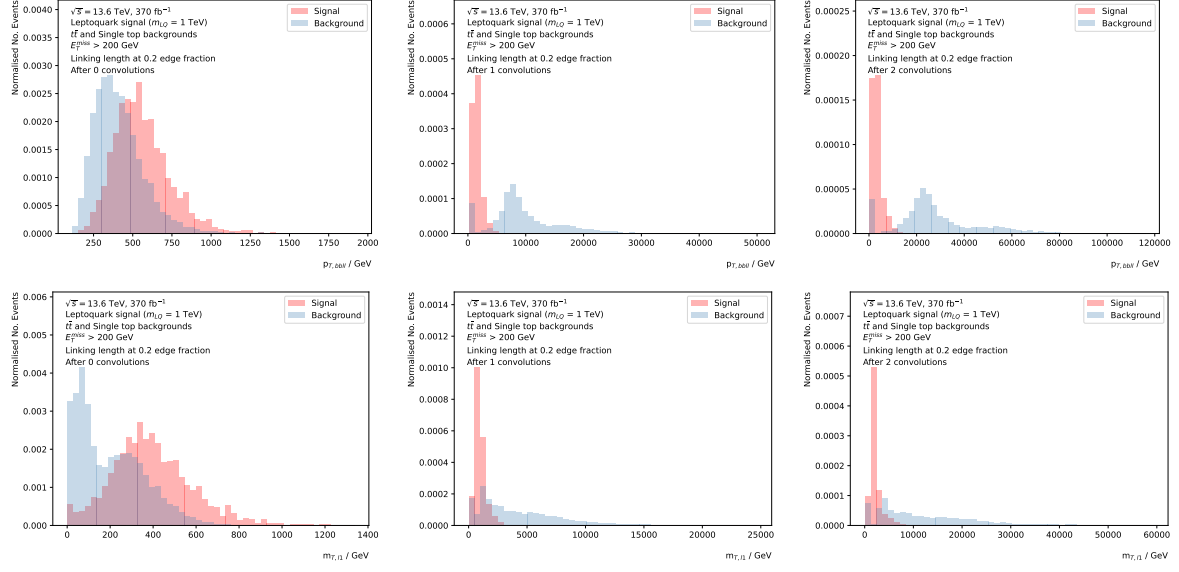


Figure G.2: Distributions of the original, once convoluted and twice convoluted $p_{T,bbll}$ and $m_{T,l1}$ in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in the high-level kinematic domain using the cosine distance with a linking length corresponding to a target edge fraction of 0.2. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

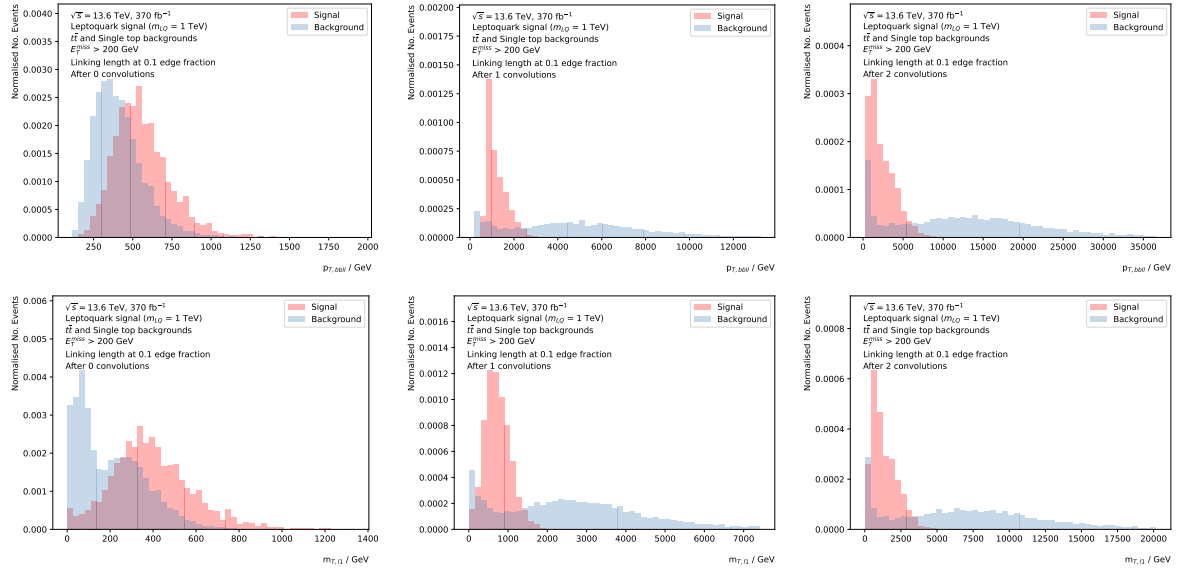


Figure G.3: Distributions of the original, once convoluted and twice convoluted $p_{T,bbll}$ and $m_{T,l1}$ in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in the high-level kinematic domain using the Earth Mover's Distance (EMD) with a linking length corresponding to a target edge fraction of 0.1. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

For the low-level kinematic variables, the $p_{T,l1}$ and E_T^{miss} are shown as two examples. Their convoluted distributions using graphs constructed in the low-level kinematic space, using the Euclidean, cosine

and the EMD distances with a linking length corresponding to a 0.1 edge fraction are shown in Figures G.4, G.5 and G.6 respectively.

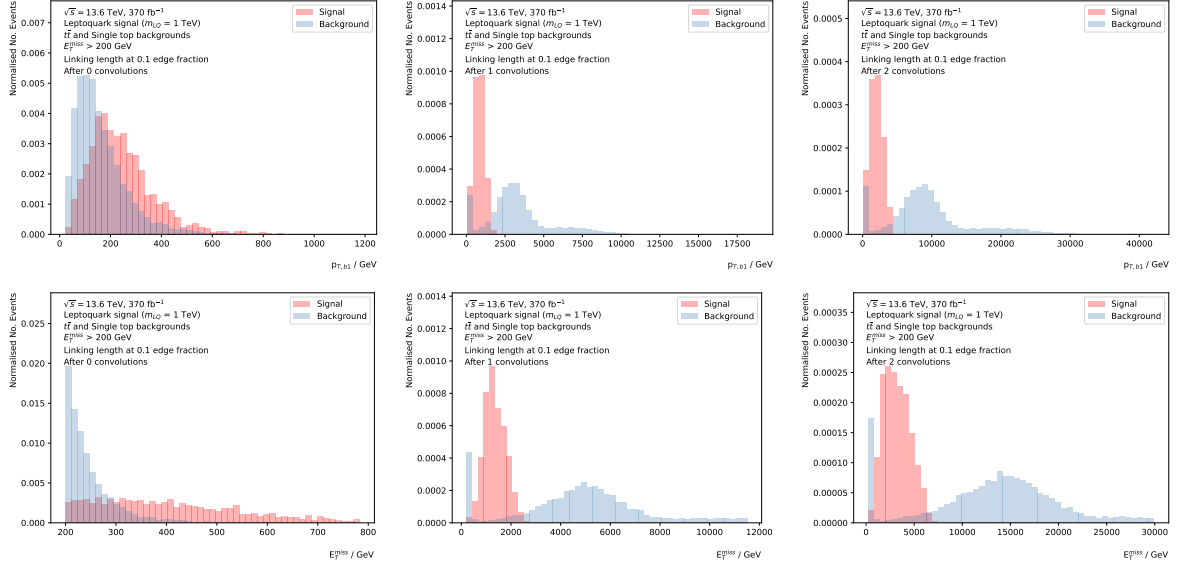


Figure G.4: Distributions of the original, once convoluted and twice convoluted $p_{T,b1}$ and E_T^{miss} in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in the low-level kinematic domain using the Euclidean distance with a linking length corresponding to a target edge fraction of 0.1. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

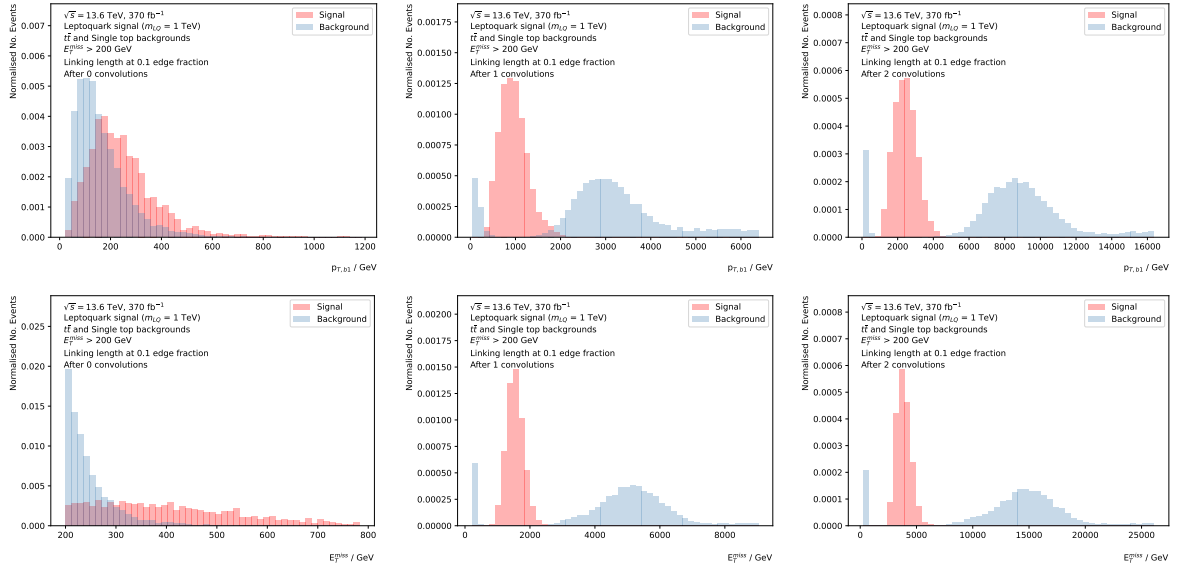


Figure G.5: Distributions of the original, once convoluted and twice convoluted $p_{T,b1}$ and E_T^{miss} in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in the low-level kinematic domain using the cosine distance with a linking length corresponding to a target edge fraction of 0.1. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

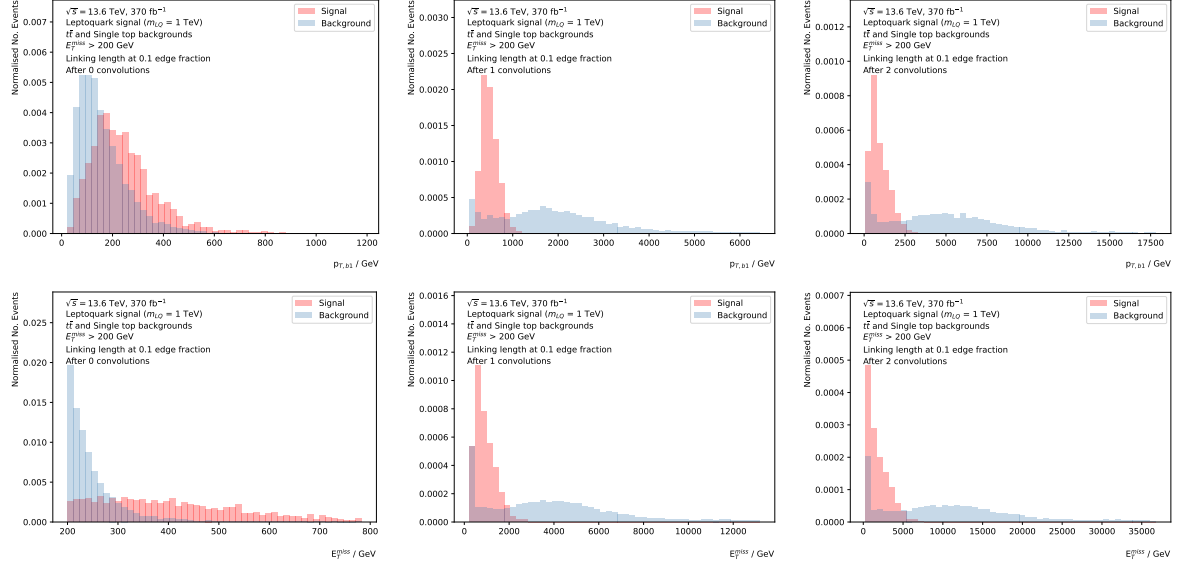


Figure G.6: Distributions of the original, once convoluted and twice convoluted $p_{T,b1}$ and E_T^{miss} in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in the low-level kinematic domain using the Earth Mover's Distance (EMD) metric with a linking length corresponding to a target edge fraction of 0.1. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

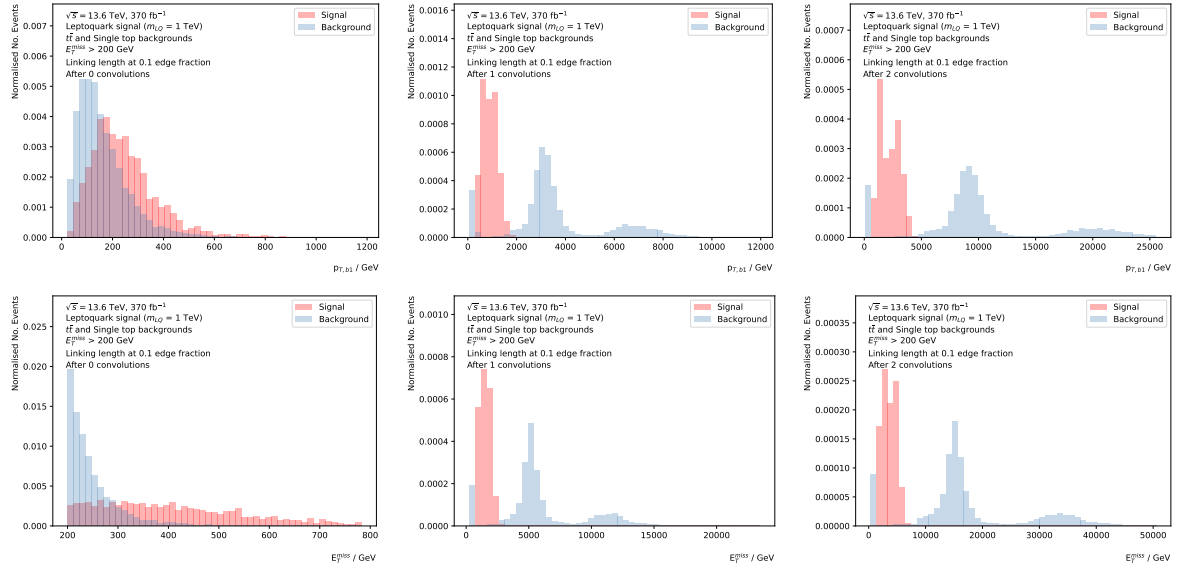


Figure G.7: Distributions of the original, once convoluted and twice convoluted $p_{T,b1}$ and E_T^{miss} in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in a 12-dimensional latent space using the Euclidean distance with a linking length corresponding to a target edge fraction of 0.1. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

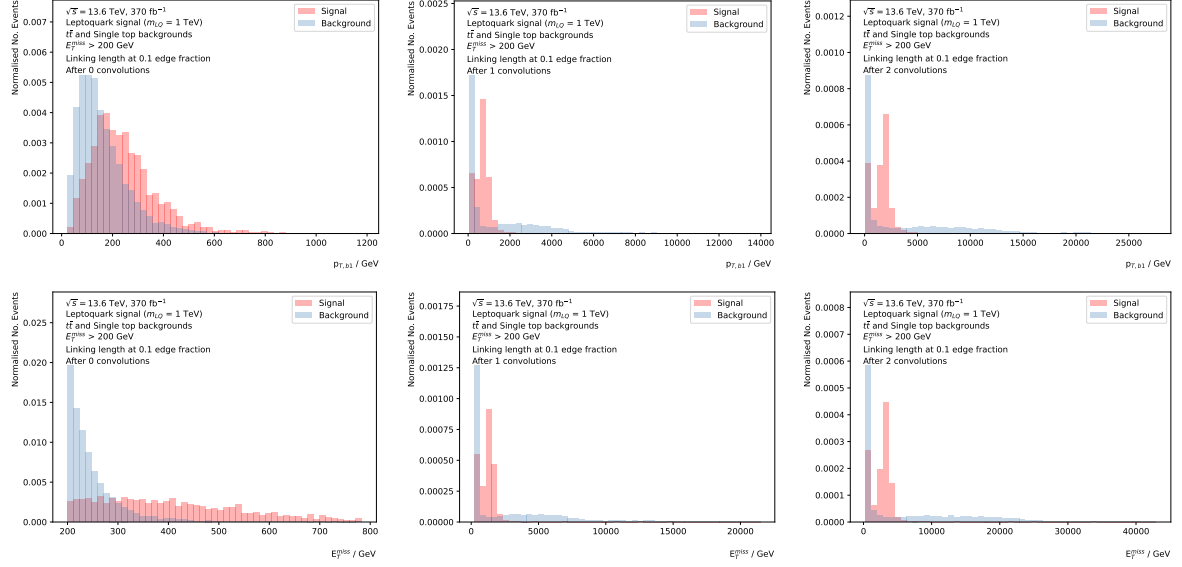


Figure G.8: Distributions of the original, once convoluted and twice convoluted $p_{T,b1}$ and E_T^{miss} in the Leptoquark signal with $m_{LQ} = 1$ TeV and the SM $t\bar{t}$ and single top backgrounds. The convolution function $\mathbf{X}' = \tilde{\mathbf{D}}^{-1/2} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{1/2} \mathbf{X}$ of the GCN is used. The graph is constructed in a 12-dimensional latent space using the cosine distance with a linking length corresponding to a target edge fraction of 0.1. Edge weights defined in Equation 6.8 are included in each iteration of convolution following Equation 6.11.

G.2 GNN Training

The training and validation losses of the GCN models trained using high-level kinematic variables as input features are shown in Figure G.9 and for the GraphConv models in Figure G.10.

The training and validation losses of the GCN models trained using low-level kinematic variables as input features, using graphs constructed in the low-level kinematic space are shown in Figure G.11 and for the GraphConv models in Figure G.12, and ones using graphs constructed in the latent space are shown in Figure G.13.

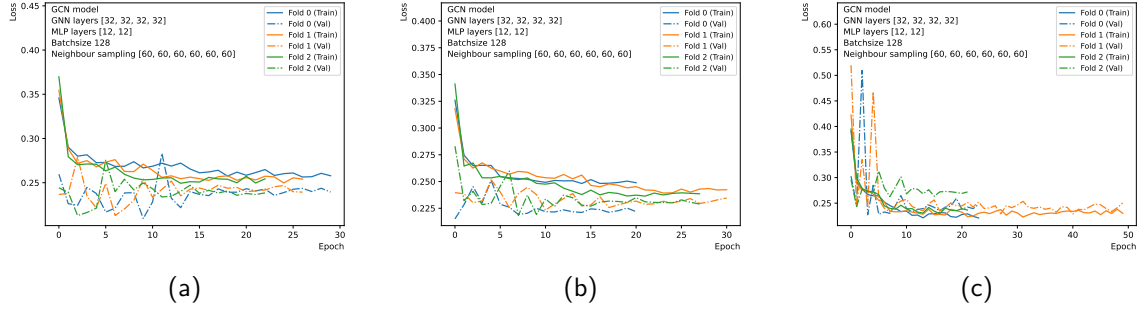


Figure G.9: Training and validation losses of GCN models trained using high-level kinematics and the Euclidean (a), cosine (b) and EMD (c) distance metrics, for each k -fold as a function of the number of epochs.

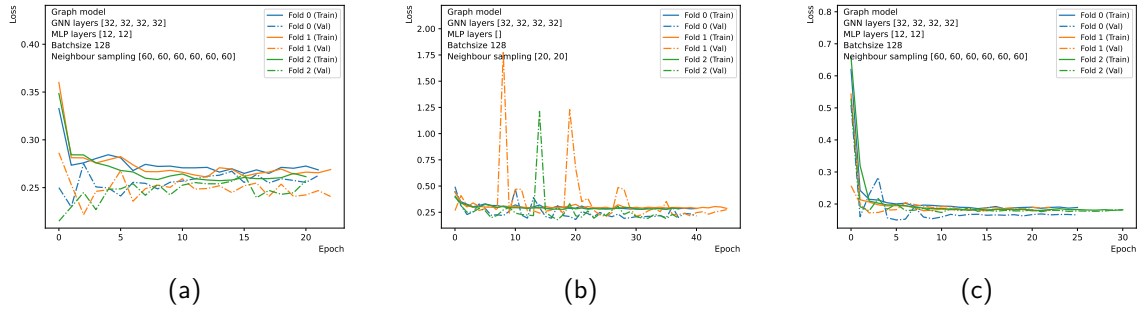


Figure G.10: Training and validation losses of GraphConv models trained using the high-level kinematics and the Euclidean (a), cosine (b) and the EMD (c) distance metrics, for each k -fold as a function of the number of epochs.

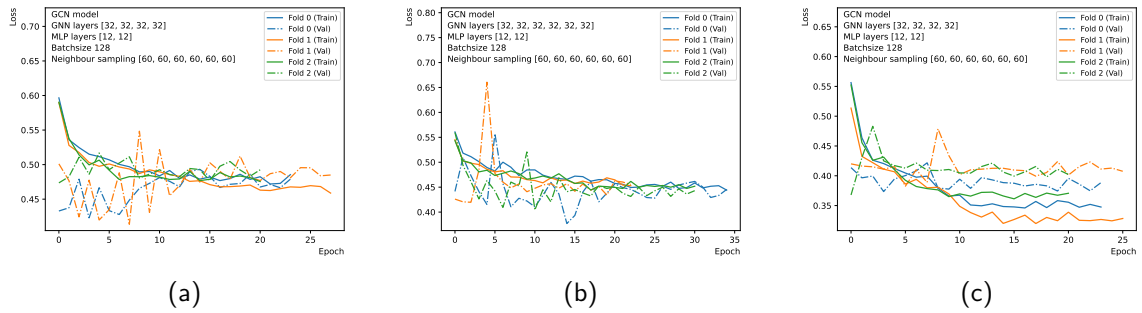


Figure G.11: Training and validation losses of GCN models trained using low-level kinematics and the Euclidean (a), cosine (b) and EMD (c) distance metrics in the low-level kinematic space, for each k -fold as a function of the number of epochs.

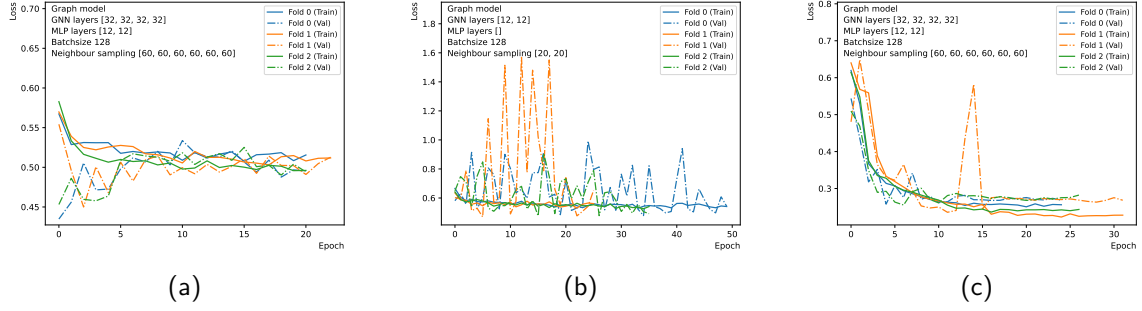


Figure G.12: Training and validation losses of GraphConv models trained using low-level kinematics and the Euclidean (a), cosine (b) and EMD (c) distance metrics in the low-level kinematic space, for each k -fold as a function of the number of epochs.

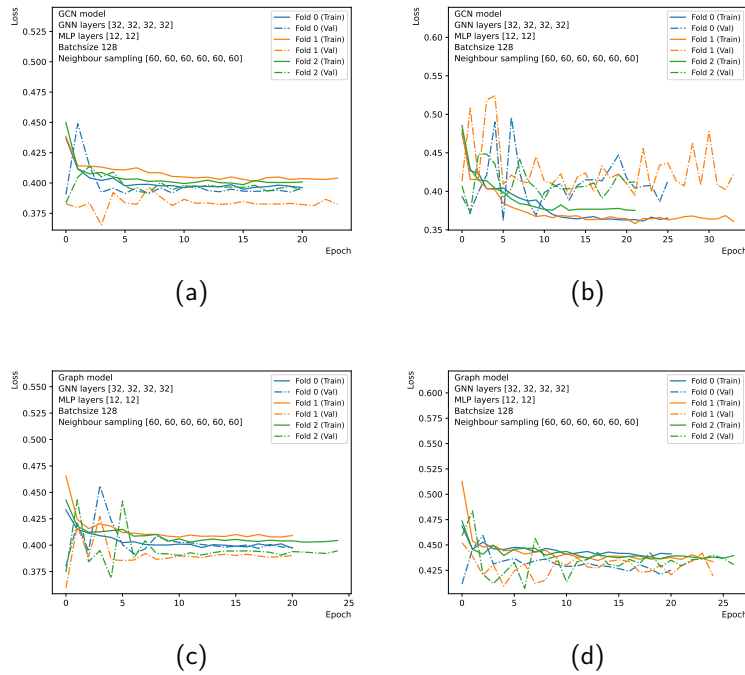


Figure G.13: Training and validation losses of GCN models trained using low-level kinematics and the Euclidean (a), cosine (b) distances, and of GraphConv models using the Euclidean (c) and cosine (d) distances. The graphs are constructed in the low-level kinematic space. The losses are plotted for each k -fold as functions of the number of epochs.

H | A Quick Study on My Thesis Writing Habits

The writing of this thesis spanned a total of 9 months, of which active writing occurred only in the last 5 months, from December 2024 to May 2025. The page count of this thesis as a function of time is shown in Figure H.1. For each entry, the colour indicates the chapter with the most number of commits made to the Gitlab repository for that day. The writing was done in a somewhat organised manner, where each chapter is worked on consecutively for the majority of the writing period. The page count increases largely linearly, with the exceptions of the initial neglect in the first 4 months, the ski season in February 2025, an impromptu climbing trip to the Peak District in April 2025, and a period of blind panic leading up to submission in May 2025.

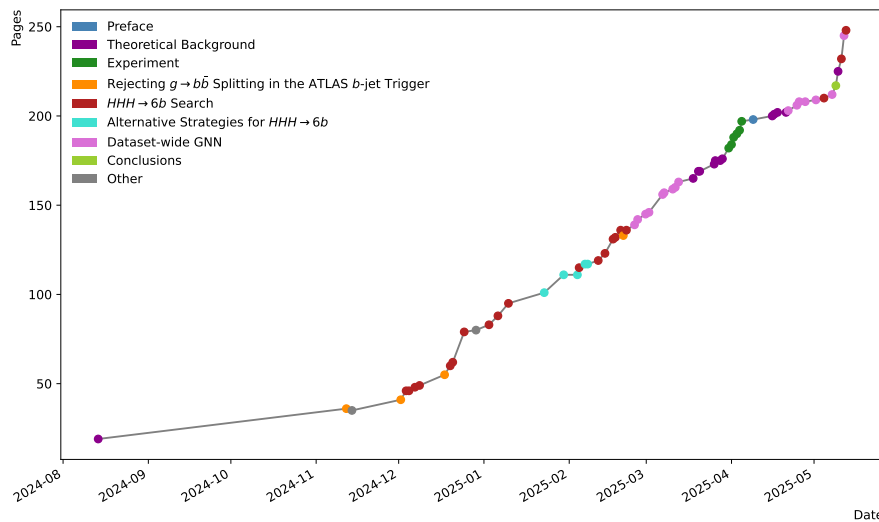


Figure H.1: Page count of the thesis as a function of time. The data points are colour coded according to the chapter with the most number of commits made to the Gitlab repository for each date entry.